

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number  
**WO 03/020916 A2**

- (51) International Patent Classification<sup>7</sup>: C12N (74) Agents: RIGAUT, Kathleen, D. et al.; Dann, Dorfman, Herrell and Skillman, 1601 Market Street, Suite 720, Philadelphia, PA 19103 (US).
- (21) International Application Number: PCT/US02/09663
- (22) International Filing Date: 28 March 2002 (28.03.2002) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/315,529 29 August 2001 (29.08.2001) US (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): UNIVERSITY OF WYOMING [US/US]; Research Products Center Office, P.O. Box 3672, Laramie, WY 82071 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): LEWIS, Randolph, V. [US/US]; 1948 Howe Road, Laramie, WY 82070 (US). HAYASHI, Cheryl, Y. [US/US]; 600 Central Avenue, No. 365, Riverside, CA 92507 (US). GATESY, John, E. [US/US]; 600 Central Avenue, No. 365, Riverside, CA 92507 (US). MOTRIUK, Dagmara [PL/US]; 1300 1/2 Custer, Laramie, WY 82070 (US).
- Published:**  
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 03/020916 A2

(54) Title: SPIDER SILK PROTEIN ENCODING NUCLEIC ACIDS, POLYPEPTIDES, ANTIBODIES AND METHOD OF USE THEREOF

(57) Abstract: Spider silk protein encoding nucleic acids, polypeptides and antibodies immunologically specific therefore are disclosed. Methods of use thereof are also provided.

**SPIDER SILK PROTEIN ENCODING NUCLEIC ACIDS, POLYPEPTIDES,  
ANTIBODIES AND METHODS OF USE THEREOF**

**By Randolph V. Lewis,**

**Cheryl Y. Hayashi,**

5

**John E. Gatesy,**

**and**

**Dagmara Motriuk**

10 Pursuant to 35 U.S.C. §202(c) it is acknowledged  
that the U.S. Government has certain rights in the  
invention described herein, which was made in part with  
funds from the National Science Foundation, Grant Number  
MCB-9806999.

15 **FIELD OF THE INVENTION**

This invention relates to the fields of molecular  
and cellular biology. Specifically, nucleic acids  
encoding spider silk polypeptides, spider silk  
polypeptides, spider silk polypeptide-specific  
20 antibodies, and methods of use thereof are provided.

**BACKGROUND OF THE INVENTION**

Several publications are referenced in this  
application by numerals in parentheses in order to more  
25 fully describe the state of the art to which this  
invention pertains. Full citations for these references  
are found at the end of the specification. The  
disclosure of each of these publications is incorporated  
by reference herein.

30 Spider silks comprise a model system for exploring  
the relationship between the amino acid composition of a  
protein, the structural properties that result from  
variations in the amino acid composition of a protein,  
and how such variations impact protein function. While

silk production has evolved multiple times within arthropods, silk use is most highly developed in spiders. Spiders are unique in their lifelong ability to spin an array of different silk proteins (or fibroin proteins) and the degree to which they depend on this ability.

There are over 34,000 described species of Araneae (1). Each species utilizes silk, and some ecribellate orb-weavers (Araneoidea) have a varied toolkit of task-specific silks with divergent mechanical properties (2).

Araneoid major ampullate silk, the primary dragline, is extremely tough. Minor ampullate silk, used in web construction, has high tensile strength. An orb-web's capture spiral, in part composed of flagelliform silk, is elastic and can triple in length before breaking (3).

Each of these fibers is composed of one or more proteins encoded by the spider silk fibroin gene family (4). Sequencing of araneoid fibroins has revealed that these fibroins are dominated by iterations of four simple amino acid motifs: poly-alanine ( $A_n$ ), alternating glycine and alanine (GA), GGX (where X represents a small subset of amino acids), and GPG(X) $_n$  (5).

Spiders draw fibers from dissolved fibroin proteins that are stored in specialized sets of abdominal glands. Each type of silk is secreted and stored by a different abdominal gland until extruded by tiny spigots on the spinnerets. Spiders use fibroin proteins singly or in combination for a variety of different purposes, including: draglines, retreats, egg sacs, and prey-catching snares. Given these specialized applications, individual silks appear to have evolved to possess mechanical properties (e.g., tensile strength and flexibility) that optimize their utility for particular applications.

Orb web spiders like *Nephila* are known to produce spider silk proteins derived from several types of silk



synthetic glands and are designated according to their organ of origin. Spider silk proteins known to exist include: major ampullate spider proteins (MaSp), minor ampullate spider proteins (MiSp), and flagelliform (Flag), tubuliform, aggregate, aciniform, and pyriform spider silk proteins. Spider silk proteins derived from each organ are generally distinguishable from those derived from other synthetic organs by virtue of their physical and chemical properties, which render them well suited to different uses. Tubuliform silk, for example, is used in the outer layers of egg-sacs, whereas aciniform silk is involved in wrapping prey and pyriform silk is laid down as the attachment disk.

Most molecular and structural investigations of spider silks have focused on dragline silk, which has an extraordinarily high tensile strength (e.g. Xu & Lewis, Proc. Natl. Acad. Sci., USA 87, 7120-7124, 1990; Hinman & Lewis, J. Biol. Chem. 267, 19320-19324, 1992; Thiel et al., Biopolymers 34, 1089-1097, 1994; Simmons et al., Science 271, 84-87, 1996; Kümmerlen et al., Macromol. 29, 2920-2928, 1996; and Osaki, Nature 384, 419, 1996). Dragline silk, often referred to as major ampullate silk because it is produced by the major ampullate glands, has a high tensile strength ( $5 \times 10^9 \text{ Nm}^{-2}$ ) similar to Kevlar ( $4 \times 10^9 \text{ Nm}^{-2}$ ) (Gosline et al., Endeavour 10, 37-43, 1986; Stauffer et al., J. Arachnol. 22, 5-11, 1994). In addition to this exceptional strength, dragline silk also exhibits substantial (~35%) elasticity (Gosline et al., Endeavour 10, 37-43, 1986). Thus a structure/function analysis of dragline silk is revealing in terms of the features of a protein which confer strength and elasticity.

Silk strength is widely attributed to crystalline beta-sheet structures. Such protein domains are found in both lepidopteran silks (e.g. Bombyx mori, Mita et al.,

J. Mol. Evol. 38, 583-592, 1994) and spider silks (Xu & Lewis, Proc. Natl. Acad. Sci., USA 87, 7120-7124, 1990; Hinman & Lewis, J. Biol. Chem. 267, 19320-19324, 1992; Gosline et al., Endeavour 10, 37-43, 1986). In contrast, elasticity is generally thought to involve amorphous regions (Wainwright et al., Mechanical design in organisms, Princeton University Press, Princeton, 1982). More precise characterization of these amorphous components can be revealed by molecular sequence data.

Based on the protein sequences of major ampullate silk proteins, a beta-turn structure was suggested to be the likely mechanism of elasticity (Hinman & Lewis, J. Biol. Chem. 267, 19320-19324, 1992). Assessing this proposition, however, was problematic because dragline silk is a hybrid of at least two distinct proteins which impart both strength and moderate elasticity.

*Nephila* minor ampullate silk can be distinguished from *Nephila* major ampullate silk by both physical and chemical properties. On a basic level, the amino acid composition of solubilized minor ampullate silk differs from that of solubilized major ampullate silk. Like the major ampullate silk proteins (major spidroin 1, MaSP1; major spidroin 2, MaSP2), the proteins comprising minor ampullate silk (minor spidroin 1, MiSP1; minor spidroin 2, MiSP2) have a primary structure dominated by imperfect repetition of a short sequence of amino acids. Moreover, in contrast to the elasticity exhibited by major ampullate silk, minor ampullate silk yields without recoil. Minor ampullate silk will stretch to about 25% of its initial length before breaking, thereby exhibiting a tensile strength of nearly 100,000 pounds per square inch (psi). The minor ampullate silk proteins, therefore, exhibit comparatively lower tensile strength and elasticity relative to major ampullate silk proteins.

The capture spiral, on the other hand, is formed

from silk proteins derived from the flagelliform and aggregate silk glands. The capture spiral of an orb-web comprises a structure having significant ability to stretch, as would be anticipated for a structure that must capture and retain prey. The capture thread has a lower tensile strength ( $1 \times 10^9 \text{ Nm}^{-2}$ ) but several times the elasticity (>200%) of dragline silk (Vollrath & Edmonds, Nature 340, 305-307, 1989; Kohler & Vollrath, J. Exp. Zool. 271, 1-17, 1995). The flagelliform silk comprises the core fiber of the spiral, while aggregate silk provides a non-fibrous, aqueous coating. Thus, while aggregate silk is an integral part of the elastic capture spiral, it is flagelliform silk that provides the ability to stretch.

#### SUMMARY OF THE INVENTION

In view of the unique properties of different silks produced by spiders, the identification of novel spider silk proteins and characterization of their chemical and physical properties provide useful new reagents having utility for a number of applications. Spider silk proteins are unique in that they possess properties which include, but are not limited to, high tensile strength and elasticity. Moreover, individual spider silk proteins have evolved to possess different combinations of properties that contribute to the physical balance between protein strength and elasticity.

Spider silk is composed of fibers formed from proteins. Naturally occurring spider silk fibers can be composites of two or more proteins. In general, spider silk proteins are found to have primary amino acid sequences that can be characterized as indirect repeats of a short consensus sequence. Variation in the consensus sequence is then responsible for the distinguishable properties of different silk proteins.

Silk fibers can be made from synthetic polypeptides having amino acid sequences substantially similar to a consensus repeat unit of a silk protein or from polypeptides expressed from nucleic acid sequences encoding a natural or engineered silk protein, or derivative thereof. Depending on the application for which a synthetic spider silk protein is intended, it may also be desirable to form fibers from a single spider silk protein or combinations of different spider silk proteins, the ratio of which can be modified accordingly.

According to one aspect of the invention, nucleic acid sequences encoding novel spider silk proteins are provided. Exemplary nucleic acid sequences of the invention have sequences comprising SEQ ID NOS: 1-28.

In a particular aspect of the invention, exemplary nucleic acid sequences encoding novel MaSp1-like spider silk proteins are provided. Exemplary nucleic acid sequences of this type have sequences comprising SEQ ID NOS: 1-7.

In another aspect of the invention, exemplary nucleic acid sequences encoding novel MaSp2-like spider silk proteins are provided. Exemplary nucleic acid sequences of this type have sequences comprising SEQ ID NOS: 8-16.

In another aspect of the invention, exemplary nucleic acid sequences encoding novel flagelliform (flag)-like spider silk proteins are provided. Exemplary nucleic acid sequences of this type have sequences comprising SEQ ID NOS: 17 and 18.

In another aspect of the invention, nucleic acid sequences encoding novel spider silk proteins are provided. Exemplary nucleic acid sequences of this type have sequences comprising SEQ ID NOS: 19 and 20.

In yet another aspect of the invention, nucleic acid sequences encoding novel spider silk proteins which

comprise atypical repetitive motifs are provided.  
Exemplary nucleic acid sequences of this type have  
sequences comprising SEQ ID NOs: 21-27.

5 In a particular aspect of the invention, an isolated  
nucleic acid sequence which encodes a novel spider silk  
protein comprising atypical repetitive motifs is  
provided. An exemplary nucleic acid sequence of this  
type has a sequence comprising SEQ ID NO: 28.

10 In a preferred embodiment of the invention, the  
isolated nucleic acid molecules provided encode spider  
silk proteins. In a particularly preferred embodiment,  
spider silk proteins of the present invention have amino  
acid sequences comprising SEQ ID NOS: 29-56.

15 In a particular aspect of the invention, novel  
MaSp1-like spider silk proteins have amino acid sequences  
comprising SEQ ID NOs: 29-35.

In another aspect of the invention, novel MaSp2-like  
spider silk proteins have amino acid sequences comprising  
SEQ ID NOs: 36-44.

20 In another aspect of the invention, novel flag-like  
spider silk proteins have amino acid sequences comprising  
SEQ ID NOs: 45 and 46.

In another aspect of the invention, novel spider  
silk proteins have amino acid sequences comprising SEQ ID  
25 NOs: 47 and 48.

In yet another aspect of the invention, novel spider  
silk proteins comprising atypical repetitive motifs have  
amino acid sequences comprising SEQ ID NOs: 49-55.

30 In a particular aspect of the invention, a novel  
spider silk protein comprising atypical repetitive motifs  
is provided. An exemplary spider silk protein amino acid  
sequence of this type comprises SEQ ID NO: 56.

According to another aspect of the present  
invention, an isolated nucleic acid molecule is provided,  
35 which has a sequence selected from the group consisting

of: (1) SEQ ID NOs: 1-28; (2) a sequence specifically hybridizing with preselected portions or all of an individual complementary strand of SEQ ID NOs: 1-28 comprising nucleic acids encoding amino acids of SEQ ID NOs: 29-56; (3) a sequence encoding preselected portions of SEQ ID NOs: 1-28, and (4) a sequence comprising nucleic acids encoding amino acids of a consensus sequence (SEQ ID NO: 57) which was derived from SEQ ID NO: 56.

Such partial sequences are useful as probes to identify and isolate homologues of spider silk protein genes of the invention. Additionally, isolated nucleic acid sequences encoding natural allelic variants of the nucleic acids of SEQ ID NOs: 1-28 are also contemplated to be within the scope of the present invention. The term natural allelic variants will be defined hereinbelow.

According to another aspect of the present invention, antibodies immunologically specific for the spider silk proteins described hereinabove are provided.

In yet another aspect of the invention, host cells comprising at least one of the spider silk protein encoding nucleic acids are provided. Such host cells include but are not limited to bacterial cells, fungal cells, insect cells, mammalian cells, and plant cells. Host cells overexpressing one or more of the spider silk protein encoding nucleic acids of the invention provide valuable reagents for many applications, including, but not limited to, production of silk fibers comprising at least one silk protein that can be incorporated into a material to modulate the structural properties of the material.

Naturally occurring spider silk proteins have an imperfectly repetitive structure. Imperfections in the repetition are likely to be a consequence of the process

by which the silk protein genes evolved, rather than a requirement for fiber formation. Imperfections in repetition are thus not likely to affect properties of fibers formed following aggregation of protein molecules.

5           Accordingly, in another embodiment of the present invention nucleic acid sequences are provided which encode engineered spider silk proteins, each of which comprises a polypeptide having direct repeats of a unit amino acid sequence. Alternatively, nucleic acid  
10           sequences may include several different unit amino acid sequences to form a "copolymer" silk protein.

          In yet another embodiment of the present invention a spider silk protein expressed from a nucleic acid sequence is provided, wherein the nucleic acid sequence  
15           is obtained from cDNA, genomic DNA, synthetic DNA, or fragments of all of the above, derived from a spider ampullate gland.

          In another embodiment of the present invention fibers made from silk protein obtained by expression of  
20           nucleic acid sequences encoding at least one spider silk protein are provided.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

25           Figure 1 shows an analysis of phylogenetic relationships of Araneae based on morphological evidence (1,27). Previously published spider fibroin sequences are from the two genera marked by white circles. Including data presented herein, fibroin sequences have been  
30           characterized for the taxa in red. Circles at internal nodes mark fossil calibration points. Extinct taxa that calibrate these nodes, *Macryphantes* (16 - black circle) and *Rosamygale* (13 - gray circle), are indicated, and higher level taxa are to the right of brackets.  
35           *Dolomedes*, *Plectreurys*, and *Euagrus* are from the families

Pisauridae, Plectreuridae, and Dipluridae, respectively.

Figure 2 shows consensus ensemble repeat units for non-araneoid spider fibroins. Single letter symbols for amino acids are used, and GGX, GA, and A<sub>n</sub> motifs are indicated in green, brown, and red, respectively. *Plectreurys* cDNA1 and *Plectreurys* cDNA2 were derived from the larger ampule-shaped glands of *Plectreurys*, and *Plectreurys* cDNA3 and *Plectreurys* cDNA4 were from the smaller ampullate glands of this spider.

Figure 3 shows consensus ensemble repeat units for four araneoid fibroin orthologue groups. Single letter symbols for amino acids are used, and GGX, GA, A<sub>n</sub>, and GPG(X)<sub>n</sub> motifs are indicated in green, brown, red, and blue, respectively. The "[spacer]" region of the MiSp fibroins is a serine-rich sequence that is 137 amino acids long in *Nephila clavipes* (Genbank #AF027735). *Nep.c.*=*Nephila clavipes*, *Nep.m.*=*N. madagascariensis*, *Nep.s.*=*N. senegalensis*, *Tet.k.*=*Tetragnatha kauaiensis*, *Tet.v.*=*T. versicolor*, *Lat.g.*=*Latrodectus geometricus*, *Arg.t.*=*Argiope trifasciata*, *Arg.a.*=*A. aurantia*, *Ara.b.*=*Araneus bicentenarius*, *Ara.d.*=*A. diadematus*, *Gas.m.*=*Gasteracantha mammosa*. \$m\$=cDNA from major ampullate glands, \$f\$=cDNA from flagelliform glands, †=PCR/genomic clone, \*=previously published sequence. The previous designations for *A. diadematus* fibroins (4) are shown in parentheses (ADF1-4).

### DETAILED DESCRIPTION OF THE INVENTION

The physical characteristics of spider silk proteins confer unparalleled mechanical properties to these fibroins and, thus, render spider silk proteins ideally suited to a variety of applications. Identification of



novel spider silk proteins as described herein,  
therefore, provides useful tools for the generation of  
natural and synthetic spider silk proteins which can be  
woven into fibers to imbue fibers comprised of such  
5 proteins with unique properties.

In a preferred embodiment of the invention, nucleic  
acid sequences encoding novel spider silk proteins have  
sequences comprising SEQ ID NOS: 1-28.

In a particularly preferred embodiment, spider silk  
10 proteins of the present invention have amino acid  
sequences comprising SEQ ID NOS: 29-56.

In yet another preferred embodiment, a consensus  
sequence derived from SEQ ID NO: 56 has amino acid  
sequences comprising SEQ ID NO: 57.

15 Other spider silk proteins have been previously  
identified, see for example U.S. patent application Nos.  
5,773,771; 5,989,894; and 5,728,810, the entire  
disclosures of which are incorporated herein by  
reference.

## 20 I. Definitions

The following definitions are provided to facilitate  
an understanding of the present invention:

With reference to nucleic acids used in the  
25 invention, the term "isolated nucleic acid" is sometimes  
employed. This term, when applied to DNA, refers to a  
DNA molecule that is separated from sequences with which  
it is immediately contiguous (in the 5' and 3'  
directions) in the naturally occurring genome of the  
30 organism from which it was derived. For example, the  
"isolated nucleic acid" may comprise a DNA molecule  
inserted into a vector, such as a plasmid or virus  
vector, or integrated into the genomic DNA of a  
procaryote or eucaryote. An "isolated nucleic acid"  
35 molecule" may also comprise a cDNA molecule. An isolated

nucleic acid molecule inserted into a vector is also sometimes referred to herein as a recombinant nucleic acid molecule.

5 With respect to RNA molecules, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e.,  
10 in cells or tissues), such that it exists in a "substantially pure" form.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two  
15 single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an  
20 oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.  
25 Appropriate conditions enabling specific hybridization of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization  
30 between nucleic acid molecules of a specified sequence homology is set forth below (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\text{Log} [\text{Na}^+] + 0.41(\% \text{ G+C}) - 0.63 (\% \text{ formamide}) - 600/\#\text{bp in duplex}$$

35

As an illustration of the above formula, using  $[Na^+] = [0.368]$  and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the  $T_m$  is 57°C. The  $T_m$  of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C.

The term "oligonucleotide," as used herein refers to primers and probes of the present invention, and is defined as a nucleic acid molecule comprised of two or more ribo- or deoxyribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on various factors and on the particular application and use of the oligonucleotide. Preferred oligonucleotides comprise 15-50 consecutive bases of SEQ ID Nos: 1-28.

The term "probe" as used herein refers to an oligonucleotide, polynucleotide or nucleic acid, either RNA or DNA, whether occurring naturally as in a purified restriction enzyme digest or produced synthetically, which is capable of annealing with or specifically hybridizing to a nucleic acid with sequences complementary to the probe. A probe may be either single-stranded or double-stranded. The exact length of the probe will depend upon many factors, including temperature, source of probe and use of the method. For example, depending on the complexity of the target sequence, the oligonucleotide probe typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The probes herein are selected to be complementary to different strands of a particular target nucleic acid sequence. This means that the probes must be sufficiently complementary so as to be able to "specifically hybridize" or anneal with their respective target strands under a set of pre-determined conditions. Therefore, the probe sequence need not reflect the exact

complementary sequence of the target. For example, a non-complementary nucleotide fragment may be attached to the 5' or 3' end of the probe, with the remainder of the probe sequence being complementary to the target strand.

5 Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the sequence of the target nucleic acid to anneal therewith specifically.

10 The term "primer" as used herein refers to an oligonucleotide, either RNA or DNA, either single-stranded or double-stranded, either derived from a biological system, generated by restriction enzyme digestion, or produced synthetically which, when placed  
15 in the proper environment, is able to functionally act as an initiator of template-dependent nucleic acid synthesis. When presented with an appropriate nucleic acid template, suitable nucleoside triphosphate precursors of nucleic acids, a polymerase enzyme,  
20 suitable cofactors and conditions such as a suitable temperature and pH, the primer may be extended at its 3' terminus by the addition of nucleotides by the action of a polymerase or similar activity to yield a primer extension product. The primer may vary in length  
25 depending on the particular conditions and requirement of the application. For example, in diagnostic applications, the oligonucleotide primer is typically 15-25 or more nucleotides in length. The primer must be of sufficient complementarity to the desired template to  
30 prime the synthesis of the desired extension product, that is, to be able to anneal with the desired template strand in a manner sufficient to provide the 3' hydroxyl moiety of the primer in appropriate juxtaposition for use in the initiation of synthesis by a polymerase or similar  
35 enzyme. It is not required that the primer sequence

represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary bases may be interspersed within the oligonucleotide primer sequence, provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the extension product.

Polymerase chain reaction (PCR) has been described in US Patents 4,683,195, 4,800,195, and 4,965,188, the entire disclosures of which are incorporated by reference herein.

Amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form may be substituted for any L-amino acid residue, provided the desired properties of the polypeptide are retained. All amino-acid residue sequences represented herein conform to the conventional left-to-right amino-terminus to carboxy-terminus orientation.

Amino acid residues are identified in the present application according to the three-letter or one-letter abbreviations in the following Table:

5

TABLE 1

	<u>Amino Acid</u>	<u>3-letter Abbreviation</u>	<u>1-letter Abbreviation</u>
	L-Alanine	Ala	A
10	L-Arginine	Arg	R
	L-Asparagine	Asn	N
	L-Aspartic Acid	Asp	D
	L-Cysteine	Cys	C
	L-Glutamine	Gln	Q
15	L-Glutamic Acid	Glu	E
	Glycine	Gly	G
	L-Histidine	His	H
	L-Isoleucine	Ile	I
	L-Leucine	Leu	L
20	L-Methionine	Met	M
	L-Phenylalanine	Phe	F
	L-Proline	Pro	P
	L-Serine	Ser	S
	L-Threonine	Thr	T
25	L-Tryptophan	Trp	W
	L-Tyrosine	Tyr	Y
	L-Valine	Val	V
	L-Lysine	Lys	K

30

The term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein that has

35 been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in

"substantially pure" form. "Isolated" is not meant to exclude artificial or synthetic mixtures with other compounds or materials, or the presence of impurities that do not interfere with the fundamental activity, and that may be present, for example, due to incomplete purification, addition of stabilizers, or compounding into, for example, immunogenic preparations or pharmaceutically acceptable preparations.

"Mature protein" or "mature polypeptide" shall mean a polypeptide possessing the sequence of the polypeptide after any processing events that normally occur to the polypeptide during the course of its genesis, such as proteolytic processing from a polyprotein precursor. In designating the sequence or boundaries of a mature protein, the first amino acid of the mature protein sequence is designated as amino acid residue 1. As used herein, any amino acid residues associated with a mature protein not naturally found associated with that protein that precedes amino acid 1 are designated amino acid -1, -2, -3 and so on. For recombinant expression systems, a methionine initiator codon is often utilized for purposes of efficient translation. This methionine residue in the resulting polypeptide, as used herein, would be positioned at -1 relative to the mature protein sequence.

A low molecular weight "peptide analog" shall mean a natural or mutant (mutated) analog of a protein, comprising a linear or discontinuous series of fragments of that protein and which may have one or more amino acids replaced with other amino acids and which has altered, enhanced or diminished biological activity when compared with the parent or nonmutated protein.

The term "biological activity" is a function or set of functions performed by a molecule in a biological context (i.e., in an organism or an in vitro surrogate or facsimile model). For spider silk proteins, biological

activity is characterized by physical properties (e.g., tensile strength and elasticity) as described herein.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, polypeptide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, mass spectrometry and the like).

The term "tag," "tag sequence" or "protein tag" refers to a chemical moiety, either a nucleotide, oligonucleotide, polynucleotide or an amino acid, peptide or protein or other chemical, that when added to another sequence, provides additional utility or confers useful properties, particularly in the detection or isolation, of that sequence. Thus, for example, a homopolymer nucleic acid sequence or a nucleic acid sequence complementary to a capture oligonucleotide may be added to a primer or probe sequence to facilitate the subsequent isolation of an extension product or hybridized product. In the case of protein tags, histidine residues (e.g., 4 to 8 consecutive histidine residues) may be added to either the amino- or carboxy-terminus of a protein to facilitate protein isolation by chelating metal chromatography. Alternatively, amino acid sequences, peptides, proteins or fusion partners representing epitopes or binding determinants reactive with specific antibody molecules or other molecules (e.g., flag epitope, c-myc epitope, transmembrane epitope of the influenza A virus hemagglutinin protein, protein A, cellulose binding



domain, calmodulin binding protein, maltose binding protein, chitin binding domain, glutathione S-transferase, and the like) may be added to proteins to facilitate protein isolation by procedures such as affinity or immunoaffinity chromatography. Chemical tag moieties include such molecules as biotin, which may be added to either nucleic acids or proteins, and facilitates isolation or detection by interaction with avidin reagents, and the like. Numerous other tag moieties are known to, and can be envisioned by the trained artisan, and are contemplated to be within the scope of this definition.

A "vector" is a replicon, such as a plasmid, cosmid, bacmid, phage or virus, to which another genetic sequence or element (either DNA or RNA) may be attached so as to bring about the replication of the attached sequence or element. An "expression vector" is a specialized vector that contains a gene with the necessary regulatory regions needed for expression in a host cell.

The term "operably linked" means that the regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to effect expression of the coding sequence. This same definition is sometimes applied to the arrangement of coding sequences and transcription control elements (e.g. promoters, enhancers, and termination elements) in an expression vector. This definition is also sometimes applied to the arrangement of nucleic acid sequences of a first and a second nucleic acid molecule wherein a hybrid nucleic acid molecule is generated.

The phrase "consisting essentially of" when referring to a particular nucleotide or amino acid means a sequence having the properties of a given SEQ ID NO:.. For example, when used in reference to an amino acid

sequence, the phrase includes the sequence per se and molecular modifications that would not affect the basic and novel characteristics of the sequence.

5 A "clone" or "clonal cell population" is a population of cells derived from a single cell or common ancestor by mitosis.

A "cell line" is a clone of a primary cell or cell population that is capable of stable growth *in vitro* for many generations.

10 An "immune response" signifies any reaction produced by an antigen, such as a viral antigen, in a host having a functioning immune system. Immune responses may be either humoral in nature, that is, involve production of immunoglobulins or antibodies, or cellular in nature,  
15 involving various types of B and T lymphocytes, dendritic cells, macrophages, antigen presenting cells and the like, or both. Immune responses may also involve the production or elaboration of various effector molecules such as cytokines, lymphokines and the like. Immune  
20 responses may be measured both in *in vitro* and in various cellular or animal systems. Such immune responses may be important in protecting the host from disease and may be used prophylactically and therapeutically.

An "antibody" or "antibody molecule" is any  
25 immunoglobulin, including antibodies and fragments thereof, that binds to a specific antigen. The term includes polyclonal, monoclonal, chimeric, and bispecific antibodies. As used herein, antibody or antibody molecule contemplates both an intact immunoglobulin molecule and  
30 an immunologically active portion of an immunoglobulin molecule such as those portions known in the art as Fab, Fab', F(ab')<sub>2</sub> and F(v).

With respect to antibodies, the term  
"immunologically specific" refers to antibodies that bind  
35 to one or more epitopes of a protein or compound of

interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

"Natural allelic variants", "mutants" and  
5 "derivatives" of particular sequences of nucleic acids refer to nucleic acid sequences that are closely related to a particular sequence but which may possess, either naturally or by design, changes in sequence or structure. By closely related, it is meant that at least about 75%,  
10 but often, more than 90%, of the nucleotides of the sequence match over the defined length of the nucleic acid sequence referred to using a specific SEQ ID NO. Changes or differences in nucleotide sequence between closely related nucleic acid sequences may represent  
15 nucleotide changes in the sequence that arise during the course of normal replication or duplication in nature of the particular nucleic acid sequence. Other changes may be specifically designed and introduced into the sequence for specific purposes, such as to change an amino acid  
20 codon or sequence in a regulatory region of the nucleic acid. Such specific changes may be made *in vitro* using a variety of mutagenesis techniques or produced in a host organism placed under particular selection conditions that induce or select for the changes. Such sequence  
25 variants generated specifically may be referred to as "mutants" or "derivatives" of the original sequence.

A "derivative" of a spider silk protein or a fragment thereof means a polypeptide modified by varying the amino acid sequence of the protein, e.g. by  
30 manipulation of the nucleic acid encoding the protein or by altering the protein itself. Such derivatives of the natural amino acid sequence may involve insertion, addition, deletion or substitution of one or more amino acids, and may or may not alter the essential activity of  
35 original the spider silk protein.

As mentioned above, the spider silk polypeptide or protein of the invention includes any analogue, fragment, derivative or mutant which is derived from a spider silk protein and which retains at least one property or other characteristic of a spider silk protein. Different "variants" of spider silk proteins exist in nature. These variants may be alleles characterized by differences in the nucleotide sequences of the gene coding for the protein, or may involve different RNA processing or post-translational modifications. The skilled person can produce variants having single or multiple amino acid substitutions, deletions, additions or replacements. These variants may include inter alia: (a) variants in which one or more amino acids residues are substituted with conservative or non-conservative amino acids, (b) variants in which one or more amino acids are added to a spider silk protein, (c) variants in which one or more amino acids include a substituent group, and (d) variants in which a spider silk protein or fragment thereof is fused with another peptide or polypeptide such as a fusion partner, a protein tag or other chemical moiety, that may confer useful properties to a spider silk protein, such as, for example, an epitope for an antibody, a polyhistidine sequence, a biotin moiety and the like. Other spider silk proteins of the invention include variants in which amino acid residues from one species are substituted for the corresponding residue in another species, either at the conserved or non-conserved positions. In another embodiment, amino acid residues at non-conserved positions are substituted with conservative or non-conservative residues. The techniques for obtaining these variants, including genetic ( suppressions, deletions, mutations, etc.), chemical, and enzymatic techniques are known to the person having ordinary skill

in the art.

To the extent such allelic variations, analogues, fragments, derivatives, mutants, and modifications, including alternative nucleic acid processing forms and alternative post-translational modification forms result  
5 in derivatives of spider silk protein that retain any of the biological properties of a spider silk protein, they are included within the scope of this invention.

The term "functional" as used herein implies that  
10 the nucleic or amino acid sequence is functional for the recited assay or purpose.

A "unit repeat" constitutes a repetitive short sequence. Thus, the primary structure of the spider silk proteins is considered to consist mostly of a series of small variations of a unit repeat. The unit repeats in  
15 the naturally occurring proteins are often distinct from each other. That is, there is little or no exact duplication of the unit repeats along the length of the protein. Synthetic spider silks, however, can be made wherein the primary structure of the protein comprises a  
20 number of exact repetitions of a single unit repeat. Additional synthetic spider silks can be synthesized which comprise a number of repetitions of one unit repeat together with a number of repetitions of a second unit  
25 repeat. Such a structure would be similar to a typical block copolymer. Unit repeats of several different sequences can also be combined to provide a synthetic spider silk protein having properties suited to a particular application.

The term "direct repeat" as used herein is a repeat  
30 in tandem (head-to-tail arrangement) with a similar repeat.

35

## **II. Preparation of Spider Silk-Encoding Nucleic Acid Molecules, Spider Silk Proteins, and Antibodies Thereto**

### **A. Nucleic Acid Molecules**

5 Nucleic acid molecules encoding the polypeptides of the invention may be prepared by two general methods: (1) synthesis from appropriate nucleotide triphosphates, or (2) isolation from biological sources. Both methods utilize protocols well known in the art. The availability of nucleotide sequence information, such as  
10 the DNA sequences encoding a spider silk protein, enables preparation of an isolated nucleic acid molecule of the invention by oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramidite method employed in the Applied Biosystems 38A DNA  
15 Synthesizer or similar devices. The resultant construct may be used directly or purified according to methods known in the art, such as high performance liquid chromatography (HPLC).

Specific probes for identifying such sequences as a  
20 spider silk protein encoding sequence may be between 15 and 40 nucleotides in length. For probes longer than those described above, the additional contiguous nucleotides are provided within sequences encoding a spider silk protein.

25 In accordance with the present invention, nucleic acids having the appropriate level of sequence homology with sequences encoding a spider silk protein may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations  
30 may be performed, according to the method of Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989), using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is  
35

carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2 X SSC and 1% SDS; (2) 15 minutes at room temperature in 2 X SSC and 0.1% SDS; (3) 5 30 minutes-1 hour at 37°C in 1 X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1 X SSC and 1% SDS, changing the solution every 30 minutes.

The nucleic acid molecules described herein include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, oligonucleotides are provided having sequences capable of hybridizing with at least one sequence of a nucleic acid sequence, such as selected segments of sequences encoding a spider silk protein. Also contemplated in the scope of the present invention are methods of use for oligonucleotide probes which specifically hybridize with DNA from sequences encoding a spider silk protein under high stringency conditions. Primers capable of specifically amplifying sequences encoding a spider silk protein are also provided. As mentioned previously, such oligonucleotides are useful as primers for detecting, isolating and amplifying sequences encoding a spider silk protein.

Antisense nucleic acid molecules which may be targeted to translation initiation sites and/or splice sites to inhibit the expression of spider silk protein genes or production of their encoded proteins are also provided. Such antisense molecules are typically between 15 and 30 nucleotides in length and often span the translational start site of a spider silk protein mRNA molecule.

#### **B. Proteins**

Full-length spider silk proteins of the present invention may be prepared in a variety of ways, according to known methods. The proteins may be purified from

appropriate sources, e.g., transformed bacterial or animal cultured cells or tissues, by immunoaffinity purification. However, this is not a preferred method due to the low levels of protein likely to be present in a given cell type at any time. The availability of nucleic acid molecules encoding spider silk proteins enables production of the proteins using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such as pSP64 or pSP65 for *in vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocytes. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or Gibco-BRL, Gaithersburg, Maryland.

Alternatively, according to a preferred embodiment, larger quantities of spider silk protein may be produced by expression in a suitable prokaryotic or eukaryotic system. For example, part or all of at least one DNA molecule, such as nucleic acid sequences having a sequence selected from the group of SEQ ID NOs: 1-28 may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*. Such vectors comprise the regulatory elements necessary for expression of the DNA in the host cell positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The spider silk proteins produced by gene expression in a recombinant prokaryotic or eukaryotic system may be purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the



recombinant protein is expressed and thereafter secreted from the host cell, to be easily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein from cell lysates (remains of cells following disruption of cellular integrity) derived from prokaryotic or eukaryotic cells in which a protein was expressed. Methods for generation of such cell lysates are known to those of skill in the art. Recombinant protein can be purified by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein or nickel columns for isolation of recombinant proteins tagged with 6-8 histidine residues at their N-terminus or C-terminus. Alternative tags may comprise the FLAG epitope or the hemagglutinin epitope. Such methods are commonly used by skilled practitioners.

The spider silk proteins of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. For example, such proteins may be subjected to amino acid sequence analysis, according to known methods.

A protein produced according to the present invention can be chemically modified after synthesis of the polypeptide. The presence of several carboxylic acid side chains (Asp or Glu) in the spacer regions facilitates the attachment of a variety of different chemical groups to silk proteins including amino acids having such side chains. The simplest and easiest procedure is to use a water-soluble carbo-diimide to attach the modifying group via a primary amine. If the group to be attached has no primary amine, a variety of linking agents can be attached via their own primary amines and then the modifying group attached via an available chemistry. Jennes, L. and Stumpf, W. E.

Neuroendocrine Peptide Methodology, Chapter 42. P.  
Michael Conn, editor. Academic Press, 1989.

Desirable chemical modifications include, but are  
not limited to, derivatization with peptides that bind to  
cells, e.g. fibroblasts, derivatization with antibiotics  
and derivatization with cross-linking agents so that  
cross-linked fibers can be made. The selection of  
derivatizing agents for a particular purpose is within  
the skill of the ordinary practitioner of the art.

#### **Exemplary Methods for Generation of Spider Silk Proteins**

In view of the unique properties of spider silk  
proteins, special considerations should be applied to the  
generation of synthetic spider silk proteins. The  
repetitive nature of amino acid sequences encoding these  
proteins may render synthesis of a full length spider  
silk protein, or fragments thereof, technically  
challenging. To facilitate production of full length  
silk protein molecules, the following protocol is  
provided.

The polypeptides of the present invention can be  
made by direct synthesis or by expression from cloned  
DNA. Means for expressing cloned DNA are set forth above  
and are generally known in the art. The following  
considerations are recommended for the design of  
expression vectors used to express DNA encoding the  
spider silk proteins of the present invention.

First, since spider silk proteins are highly  
repetitive in their structure, cloned DNA should be  
propagated and expressed in host cell strains that can  
maintain repetitive sequences in extrachromosomal  
elements (e.g. SURE™ cells, Stratagene). The prevalence  
of specific amino acids (e.g., alanine, glycine, proline,  
and glutamine) also suggests that it might be  
advantageous to use a host cell that overexpresses tRNA

for these amino acids.

The proteins of the present invention can otherwise be expressed using vectors providing for high level transcription, fusion proteins allowing affinity  
5 purification through an epitope tag, and the like. The hosts can be either bacterial or eukaryotic cells. Eukaryotic cells such as yeast, especially *Saccharomyces cerevisiae*, or insect cells might be particularly useful eukaryotic hosts. Expression of an engineered minor  
10 ampullate silk protein is described in U.S. Pat. No. 5,756,677, herein incorporated by reference. Such an approach can be used to express proteins of the present invention.

A useful spider silk protein or fragment thereof may  
15 be (1) insoluble inside a cell in which it is expressed or (2) capable of being formed into an insoluble fiber under normal conditions by which fibers are made. Preferably, the protein is insoluble under conditions (1) and (2). Specifically, the protein or fragment may be  
20 insoluble in a solvent such as water, alcohol (methanol, ethanol, etc.), acetone and/or organic acids, etc. The spider silk protein or fragment thereof should be capable of being formed into a fiber having high tensile strength, e.g., a tensile strength of 0.5x to 2x wherein  
25 x is the tensile strength of a fiber formed from a corresponding natural silk or whole protein. A spider silk protein or fragment thereof should also be capable of being formed into a fiber possessing high elasticity, e.g., at least 15%, more preferably about 25%.

30 Variants of a spider silk protein may be formed into a fiber having a tensile strength and/or elasticity which is greater than that of the natural spider silk or natural protein. The elasticity may be increased up to 100%. Variants may also possess properties of protein  
35 fragments.

A fragment or variant may have substantially the same characteristics as a natural spider silk. The natural protein may be particularly insoluble when in fiber form and resistant to degradation by most enzymes.

5           Recombinant spider silk proteins may be recovered from cultures by lysing cells to release spider silk proteins expressed therein. Initially, cell debris can be separated by centrifugation. Clarified cell lysate comprised of debris and supernatant can then be  
10           repeatedly extracted with solvents in which spider silk proteins are insoluble, but cellular debris is soluble. A differential solubilization process such as described above can be used to facilitate isolation of a purified spider silk protein precipitate. These procedures can be  
15           repeated and combined with other procedures including filtration, dialysis and/or chromatography to obtain a pure product.

          Fibrillar aggregates will form from solutions by spontaneous self-assembly of spider silk proteins when  
20           the protein concentration exceeds a critical value. The aggregates can be gathered and mechanically spun into macroscopic fibers according to the method of O'Brien et al. [I. O'Brien et al., "Design, Synthesis and  
Fabrication of Novel Self-Assembling Fibrillar Proteins",  
25           in Silk Polymers: Materials Science and Biotechnology, pp. 104-117, Kaplan, Adams, Farmer and Viney, eds., c. 1994 by American Chemical Society, Washington, D.C.].

### **Exemplary Methods for Preparation of Fibers From Spider 30           Silk Proteins**

          As noted above, the spider silk proteins can be viewed as derivatized polyamides. Accordingly, methods for producing fiber from soluble spider silk proteins are similar to those used to produce typical polyamide  
35           fibers, e.g. nylons, and the like.

O'Brien et al. *supra* describe fiber production from adenovirus fiber proteins. In a typical fiber production, spider silk proteins can be solubilized in a strongly polar solvent. The protein concentration of such a protein solution should typically be greater than 5% and is preferably between 8 and 20%.

Fibers should preferably be spun from solutions having properties characteristic of a liquid crystal phase. The fiber concentration at which phase transition can occur is dependent on the polypeptide composition of a protein or combination of proteins present in the solution. Phase transition, however, can be detected by monitoring the clarity and birefringence of the solution. Onset of a liquid crystal phase can be detected when the solution acquires a translucent appearance and registers birefringence when viewed through crossed polarizing filters.

The solvent used to dissolve a spider silk protein should be polar, and is preferably highly polar. Such solvents are exemplified by di- and tri- haloacetic acids, and haloalcohols (e.g. hexafluoroisopropanol). In some instances, co-solvents such as acetone are useful. Solutions of chaotropic agents, such as lithium thiocyanate, guanidine thiocyanate or urea can also be used.

In one fiber-forming technique, fibers can first be extruded from the protein solution through an orifice into methanol, until a length sufficient to be picked up by a mechanical means is produced. Then a fiber can be pulled by such mechanical means through a methanol solution, collected, and dried. Methods for drawing fibers are considered well-known in the art. For example, fibers made from a 58 kDa synthetic MaSp consensus polypeptide were drawn by methods similar to those used for drawing low molecular weight nylons. Such

methods are described in U.S. Pat. No. 5,994,099, the entirety of which is incorporated herein by reference.

Of note, spider silk proteins of the present invention have primary structures dominated by imperfect repetition of a short sequence of amino acids. A "unit repeat" constitutes one such short sequence. Thus, the primary structure of a spider silk protein can be thought to consist mostly of a series of small variations of a unit repeat. Unit repeats in a naturally occurring protein are often distinct from each other. In other words, there is little or no exact duplication of a unit repeat along the length of a protein. Synthetic spider silks, however, can be generated wherein the primary structure of a synthetic spider silk protein can be described as a number of exact repetitions of a single unit repeat. Additional synthetic spider silks can be described as a number of repetitions of one unit repeat together with a number of repetitions of a second unit repeat. Such a structure would be similar to a typical block copolymer. The present invention also encompasses generation of synthetic spider silk proteins comprising unit repeats derived from several different spider silk sequences (naturally occurring variants or genetically engineered variants thereof).

Such synthetic hybrid spider silk proteins may each have 900 to 2700 amino acids with 25 to 100, preferably 30 to 90 repeats. A spider silk or fragment or variant thereof usually has a molecular weight of at least about 16,000 daltons, preferably 16,000 to 150,000 daltons, more preferably 50,000 to 120,000 daltons for fragments and greater than 100,000 but less than 500,000 daltons, preferably 120,000 to 350,000 for a full length protein.

### **C. Antibodies**

The present invention also provides antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward a spider silk protein may be prepared according to standard methods. In a preferred embodiment, monoclonal antibodies are prepared, which react immunospecifically with various epitopes of the spider silk proteins described herein. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies that immunospecifically interact with spider silk proteins can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of anti-spider silk protein antibodies are described below.

### **III. Uses of Spider Silk-Encoding Nucleic Acids, Spider Silk Proteins and Antibodies Thereto**

#### **A. Spider Silk-Encoding Nucleic Acids**

Spider silk protein-encoding nucleic acids may be used for a variety of purposes in accordance with the present invention. Spider silk protein-encoding DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of genes encoding spider silk proteins. Methods in which spider silk protein-encoding nucleic acids may be utilized as probes for such assays include, but are not limited to: (1) *in situ* hybridization; (2) Southern hybridization; (3) northern hybridization; and (4) assorted amplification

reactions such as polymerase chain reactions (PCR).

5 The spider silk protein-encoding nucleic acids of the invention may also be utilized as probes to identify related genes from other animal species. As is well known in the art, hybridization stringencies may be adjusted to allow hybridization of nucleic acid probes with complementary sequences of varying degrees of homology. Thus, spider silk protein-encoding nucleic acids may be used to advantage to identify and  
10 characterize other genes of varying degrees of relation to the spider silk protein genes of the invention. Such information enables further characterization of nucleic acid sequences which encode proteins that possess physical properties typical of spider silk proteins and thus facilitate structure/function analysis of such  
15 proteins. Additionally, they may be used to identify genes encoding proteins that interact with spider silk proteins (e.g., by the "interaction trap" technique), which should further accelerate identification of other  
20 components utilized in webs comprised of spider silk proteins. Moreover, interacting proteins identified in such screens maybe of utility in the generation and/or optimization of materials comprised of synthetic spider silk proteins. Spider silk protein encoding nucleic  
25 acids may also be used to generate primer sets suitable for PCR amplification of target spider silk protein DNA. Criteria for selecting suitable primers are well known to those of ordinary skill in the art.

30 Host cells comprising at least one spider silk protein encoding DNA molecule are encompassed in the present invention. Host cells contemplated for use in the present invention include but are not limited to bacterial cells, fungal cells, insect cells, mammalian cells, and plant cells. The spider silk protein encoding  
35 DNA molecules may be introduced singly into such host



cells or in combination to assess the phenotype of cells conferred by such expression. Methods for introducing DNA molecules are also well known to those of ordinary skill in the art. Such methods are set forth in Ausubel et al. eds., Current Protocols in Molecular Biology, John Wiley & Sons, NY, NY 1995, the disclosure of which is incorporated by reference herein.

As described above, spider silk protein-encoding nucleic acids are also used to advantage to produce large quantities of substantially pure spider silk proteins, or selected portions thereof.

### B. Proteins and Antibodies

Purified spider silk protein, or fragments thereof, produced by methods of the present invention can be used to advantage in a variety of different applications, including, but not limited to, production of fabric, sutures, medical coverings, high-tech clothing, rope, reinforced plastics, and other applications in which various combinations of strength and elasticity are required.

Table II lists physical properties of various biological and manmade materials

Material	Material Strength (N m <sup>-2</sup> )	Elasticity (%)	Energy to Break (J kg <sup>-1</sup> )
Dragline Silk	4 x 10 <sup>9</sup>	35	1 x 10 <sup>5</sup>
Minor Silk	1 x 10 <sup>9</sup>	5	3 x 10 <sup>4</sup>
Flagelliform Silk	1 x 10 <sup>9</sup>	200+	1 x 10 <sup>5</sup>
KEVLAR	4 x 10 <sup>9</sup>	5	3 x 10 <sup>4</sup>
Rubber	1 x 10 <sup>9</sup>	600	8 x 10 <sup>4</sup>
Tendon	1 x 10 <sup>9</sup>	5	5 x 10 <sup>3</sup>

As shown in Table II, spider silks are characterized by advantageous physical properties, including, but not limited to, high tensile strength and pronounced elasticity, that are highly desirable for numerous applications. It is significant to note that spider silks possess these physical properties in aggregation which renders them unique proteins having unparalleled utility. For example, spider dragline silk has a tensile strength greater than steel or carbon fibers (200 ksi), elasticity as great as some nylon (35%), a stiffness as low as silk (0.6 msi), and the ability to supercontract in water (up to 60% decrease in length). In view of its high tensile strength and elasticity, the energy required to break dragline silk exceeds that required to break any known fiber including Kelvar™ and steel. These properties are unmatched by any known natural or manmade material. Moreover, the new materials of the present invention would also provide unique combinations of such desirable features in a very low weight material.

In view of the foregoing advantageous properties, use of the spider silk proteins disclosed in the present invention as components in materials would produce superior products. When spider silk is dissolved in an appropriate solvent and forced through a small orifice to generate spider silk fibers, such fibers can be woven into a fabric/material or added into a composite fabric/material. For example, spider silk fibers can be woven into fabrics to modulate the strength and elasticity of a fabric, thus rendering materials comprising such modified fabric optimized for different applications. Spider silk fibers can be of particular utility when incorporated into materials used to make high-tech clothing, rope, sails, parachutes, wings on aerial devices (e.g., hang gliders), flexible tie downs for electrical components, sutures, and even as a

biomaterial for implantation (e.g., artificial ligaments or aortic banding). Biomedical applications involve use of natural and/or synthetic spider silk fibers of the present invention in sutures used in surgical procedures, including, but not limited to: eye surgery, reconstructive surgery (e.g., nerve or tympanic membrane reconstruction), vascular closure, bowel surgery, cosmetic surgery, and central nervous system surgery. Natural and synthetic spider silk fibers may also be of utility in the generation of antibiotic impregnated sutures and implant material and matrix material for reconstruction of bone and connective tissue. Implants and matrix material for reconstruction may be impregnated with aggregated growth factors, differentiation factors, and/or cell attractants to facilitate incorporation of the exogenous material and optimize recovery of a patient. Spider silk proteins and fibers of the present invention can be used for any application in which various combinations of strength and elasticity are required. Moreover, spider silk proteins can be modified to optimize their utility in any application. As described above, sequences of spider silk proteins can be modified to alter various physical properties of a fibroin and different spider silk proteins and variants thereof can be woven in combination to produce fibers comprised of at least one spider silk protein or variant thereof.

In a preliminary study designed to evaluate the potential for an immune response to a natural spider silk protein, natural dragline silk was implanted into mice and rats intramuscularly, intraperitoneally, or subcutaneously. Animals into which natural dragline silk was introduced did not mount an immune response to the spider silk protein, irrespective of the site of implantation. Of note, tissue sections surrounding

spider silk protein implants were essentially identical to tissue sections derived from implantation sites into which a polyethylene rod was inserted. Since a polyethylene rod was used as the solid matrix about which the dragline spider silk protein was wrapped prior to implantation, introduction of a polyethylene rod alone serves as a negative control for the experiment. In view of the above, spider silk proteins of the present invention are expected to elicit minimal immunological responses when introduced into vertebrate animals.

Synthetic spider silk fibers are of utility in any application for which natural spider silk fibers can be used. For example, synthetic fibers may be mixed with various plastics and/or resins to prepare a fiber-reinforced plastic and/or resin product. Because spider silk is stable up to 180°C., spider silk protein fibers would be of utility as structural reinforcement material in thermal injected plastics.

It should be apparent from the foregoing that the spider silk proteins of the present invention and derivatives thereof can be generated in large quantities by means generally known to those of skill in the art. Spider silk proteins and derivatives thereof can be made into fibers for any intended use. Moreover, mixed composites of fibers are also of interest as a consequence of their unique combined properties. Such mixed composites can confer characteristics of flexibility and strength to any material into which they can be incorporated.

Purified spider silk protein, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which also may serve as sensitive detection reagents for the presence and accumulation of a spider silk protein (or complexes containing spider silk protein) in cells. Recombinant techniques enable

expression of fusion proteins containing part or all of a spider silk protein. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal antibodies specific for various epitopes of a spider silk protein, thereby providing even greater sensitivity for detection of a spider silk protein in cells.

Polyclonal or monoclonal antibodies immunologically specific for a spider silk protein may be used in a variety of assays designed to detect and quantitate these proteins. Such assays include, but are not limited to: (1) flow cytometric analysis; and (2) immunoblot analysis (e.g., dot blot, Western blot) of extracts from various cells. Additionally, anti-spider silk protein antibodies can be used for purification of a spider silk protein and any associated subunits (e.g., affinity column purification, immunoprecipitation).

From the foregoing discussion, it can be seen that spider silk-encoding nucleic acids, spider silk expressing vectors, spider silk and anti-spider silk antibodies of the invention can be used separately or in combination, for example, 1) to identify nucleic acid sequences encoding other spider silk proteins or proteins comprising similar motifs, 2) generate novel hybrid spider silk proteins selected for optimization of different physical properties, 3) to express large quantities of spider silk proteins or fragments or derivatives thereof, and 4) to detect expression of spider silk proteins in cells and/or organisms.

The following examples are provided to illustrate an embodiment of the invention. They are not intended to limit the scope of the invention in any way.

**EXAMPLE I****Identification of Clones Encoding Spider Silk Proteins**

5 In order to identify novel spider silk proteins and expand the limited database of nucleic acid sequences encoding fibroin proteins, eleven cDNA libraries derived from silk glands of seven spider genera were generated. cDNA data were supplemented by information from two genomic libraries and PCR-amplified sequences (7).  
10 Partial cDNA or gene sequences for 28 fibroins from seven families of Araneae were identified. The data, as described herein, greatly extend the phylogenetic diversity of characterized fibroins (Fig. 1).

**15 Methods for Collection of Sequence Data**

cDNA libraries were made from major ampullate glands of *Argiope trifasciata* (Araneidae) and *Lactrodectus geometricus* (Theridiidae), flagelliform glands of *A. trifasciata*, ampullate glands of *Dolomedes tenebrosus*  
20 (*Pisauridae*), two sets of silk glands from *Plectreurys tristis* (*Plectreuridae*), and silk glands of the mygalomorph *Euagrus chisoseus* (*Dipluridae*). Glands from *Dolomedes* were the four pairs of spindly ampullate glands with long tails that are connected to the spinnerets via  
25 extensive looped ducts. Glands from *Plectreurys* were the two largest pairs of ampule-shaped glands. The relatively uniform silk glands of *Euagrus* were combined in the RNA extraction for this species. Genomic libraries were constructed for *Nephila madagascariensis*  
30 (*Tetragnathidae*) and for *A. trifasciata*. Data from the thirteen libraries were augmented by PCR amplified genomic sequences from eight araneoids.

All the silk glands from *Phidippus audax* (*Salticidae*) were combined in the RNA extraction for this

species. Similarly, all the silk glands from *Zorocrates* sp. (Zorocratidae) were combined in the RNA extraction for this species. Separate cDNA libraries were made from the aciniform glands connected to the median spinnerets and aciniform glands connected to the posterior spinnerets of *Argiope trifasciata* (Araneidae). Identical cDNA sequences for an aciniform fibroin were isolated from both libraries.

Procedures for construction and screening of the seven cDNA libraries were as follows. Silk glands were dissected from euthanized spiders and flash-frozen in liquid nitrogen. mRNA was extracted from the glands using Dynabeads Oligo(dT)25 (Dyna). cDNA was synthesized using the SuperScript Choice System (Life Technologies) with oligo(dT) as the first-strand synthesis primer. Size-fractionated cDNAs (ChromaSpin-1000, Clontech) were ligated into either pGEM-3zf(+) (Promega) and electroporated into SURE cells (Stratagene), pZErO<sup>®</sup>-2 cells (Invitrogen) or TOP10 cells (Invitrogen). Eleven libraries of ~1500 recombinant colonies each were constructed, and colonies were replicated onto nylon membranes for screening.

Silk cDNA clones were identified by sequential hybridizations (QuikHyb, Stratagene) with the  $\gamma^{32}\text{P}$ -labeled probes CCWAYWCCNCCATATCCWCC (SEQ ID NO: 58), CCWCCWGGWCCNNNWCCWCCWGGWCC (SEQ ID NO: 59), CCWGGWCCTTGTTGWCCWGGWCC (SEQ ID NO: 60), GCDGCDGCDGCDGCDGC (SEQ ID NO: 61), CCWGCWCCWGCWCCWGCWCC (SEQ ID NO: 62), and CCAGADAGACCAGGATTACT (SEQ ID NO: 63) and through the sequencing of clones selected for large insert sizes. The above oligonucleotides were designed based on published spider silk sequences (see 4, 5, 8, 9). Hundreds of positive clones were screened with restriction enzymes, and a subset of clones was sequenced (ABI) using standard M13 sequencing primers and by

inserting transposons with the Genome Priming System (NEB). Divergent transcripts were recognized as members of the spider silk fibroin gene family by internal repetitiveness, sequence similarity to previously sequenced araneoid fibroins in the nonrepetitive COOH-terminus, and consistency of sequence translations with amino acid compositions from dissected silk glands and from published studies (J. Palmer (1985) J. Morphol. 186:195).

Genomic libraries were constructed for the tetragnathid, *Nephila madagascariensis* ( $\lambda$ Gem-12, Promega), and for the araneid *Argiope trifasciata* ( $\lambda$ FixII, Stratagene). The genomic libraries were screened with the radiolabeled probes

CCWCCWGGWCCNNNWCCWCCWGGWCC (SEQ ID NO: 59) and CCWGGWCCTTGTTGWCCWGGWCC (SEQ ID NO: 60). Selected silk gene inserts were excised from the  $\lambda$  arms, subcloned into pGEM (Promega) vectors, and sequenced as above.

Genomic sequences were amplified from the araneoids, *N. madagascariensis*, *N. senegalensis*, *A. trifasciata*, *A. aurantia*, *Tetragnatha kawaiensis*, *T. versicolor*, *Latrodectus geometricus*, and *Gasteracantha mammosa*. PCR was by standard procedures (Gibco recombinant Taq polymerase) using primers GGTGCTGGACAAGGAGGATACG (SEQ ID NO: 64), GGCTTGATAAACTGATTGACCAACG (SEQ ID NO: 65), and CACAGCCAGAGAGACCAGGATTGC (SEQ ID NO: 66) for MaSp1 and CCAGGAGGATATGGACCAGGTC (SEQ ID NO: 67), CCGACAACCTGGGCGAACTGAG (SEQ ID NO: 68), CAAGGATCTGGACAGCAAGG (SEQ ID NO: 69), CAACAAGGACCAGGAAGTGGC (SEQ ID NO: 70), CCAACCAWTTGCGCATACTG (SEQ ID NO: 71), GCTTGAGTTAAAGAYTGACC (SEQ ID NO: 72), and GCAGGACCAGGAAGTTATG (SEQ ID NO: 73) for MaSp2. A PCR reaction generally resulted in production of a ladder of DNA fragments. Such ladders result from annealing of PCR



primers to multiple binding sites in the repetitive sequences of a silk gene. For each fibroin amplified, the largest tight band in the PCR ladder was excised and cloned using the TOPO XL PCR cloning kit (Invitrogen).  
5 Two to three clones of each silk gene were sequenced as above. Additional sequences from 11 spider fibroins were taken from GenBank (accession numbers M37137, U03848, M92913, AF027735, AF027736, AF027737, AF027972, AF027973, AF218623, AF218624, U20328, U47853, U47854, U47855, and  
10 U47856). These published sequences are from the araneoid genera, *Nephila* and *Araneus* (Fig. 1).

### Results

Like previously published fibroins from spiders (4-  
15 5,8-9) and lepidopterans (10-11), the sequences of the invention encode repetitive alanine and glycine-rich proteins. In each molecule, iterated amino acid motifs are organized into higher-order ensemble repeats. Ensemble repeats within each fibroin were aligned, and a  
20 consensus ensemble repeat was generated for each molecule (12). In part, silk DNA sequences from non-araneoid spiders (Fig. 2) reiterate the importance of amino acid motifs that comprise orb-weaver fibroins. GA, GGX, and  $A_n$  form the consensus ensemble repeat units of silk  
25 fibroins from the pisaurid fishing spider, *Dolomedes*. The association of these three motifs in *Dolomedes* silk proteins mirrors the pattern seen in major and minor ampullate fibroins of orb-weavers (Fig. 3). GA, GGX, and  $A_n$  motifs are also distributed, sometimes sparsely, among  
30 ensemble repeat units from successively more basal lineages of spiders (Haplogynae and Mygalomorphae).  $A_n$  is represented in each of the fibroins from these taxa and from all lineages of Araneae studied thus far (Figs. 2 and 3). Mygalomorphae, tarantulas and their kin,  
35 diverged from Araneomorphae, "true" spiders, minimally

240 million years ago in the middle Triassic (13, Fig. 1), thus  $A_n$  motifs may have been maintained in different spider silks since that time.

Although the fibroins of *Plectreurys* (Haplogynae) and *Euagrus* (Mygalomorphae) are internally repetitive, the ensemble repeats from these basal taxa (Fig. 2) are unlike analogous units from previously described silks (Fig. 3). Each of the fibroins from these primitive groups contains stretches of serine. *Plectreurys* cDNA1 is highly internally repetitive with iterations of  $A_n$ ,  $S_n$ ,  $(GX)_n$ , and  $(AQ)_n$ . *Plectreurys* cDNA3 has a unique molecular architecture with the 5' end encoding a tandem array of long repeat units, and the 3' end encoding 15 repeats of a much shorter ensemble unit. The ~346 amino acid *Euagrus* repeat unit is a complex mixture of serine and alanine-rich sequences that includes a string of threonine, an amino acid that is rare in araneoid fibroins (Fig. 2).

Aside from an overall modular structure, scattered GA, GGX, and  $A_n$  motifs, and amino acid matches in the non-repetitive carboxy terminus, there is only limited sequence similarity between araneoid fibroins and those from *Plectreurys* and *Euagrus* (Figs. 2 and 3). Data presented herein clearly indicate that spiders utilize a broad diversity of fibroin sequences to spin silk threads, such diversity may be a reflection of the divergent ecosystems inhabited by these species (1,14). The novel fibroin repeats of basal Araneae suggest that spider silk design may not be especially dependent on specific sequences, but comparisons of fibroins among orb-weavers contradict this notion (Fig. 3).

In combination with published data (4-5,8-9), these new sequences allow comparisons between the two basal-most clades of ecribellate orb-weavers, Araneidae and "derived araneoids" (Fig. 1), for four groups of fibroins

(15): major ampullate spidroin 1-like (MaSp1), major ampullate spidroin 2-like (MaSp2), minor ampullate spidroins (MiSp), and flagelliform silk protein (Flag). Differences among fibroins within each of these four groups are primarily variations in the arrangement and frequency of  $A_n$ , GA, GGX, and GPG(X) $_n$  motifs (Fig. 3).  $A_n$ , GA, and GGX are present in consensus repeats for both araneid and derived araneoid MiSp orthologues. Major ampullate fibroins are similarly conserved among araneoids. Stable repeats for MaSp1 are  $A_n$ , GA, and GGX, and for MaSp2 are GPG(X) $_n$  and  $A_n$ . These motifs are retained even in major ampullate fibroins of the widow *Latrodectus*, a cob-web weaving araneoid that does not spin a conventional orb web. The long Flag repeats are divergent within Araneioidea, but both araneid and derived araneoid repeat units are comprised primarily of clustered GPG(X) $_n$  and GGX motifs (Fig. 3).

Fossil evidence suggests that the divergence of Araneidae from derived araneoids occurred no later than the early Cretaceous (Fig. 1). Therefore, the motifs conserved within MaSp1, MaSp2, MiSp, and Flag have been maintained, presumably by stabilizing selection, for over 125 million years (16). Motifs that have been retained over such long evolutionary periods are likely to be critical to the divergent mechanical properties of the specialized orb-weaver silks.

## EXAMPLE II

### Clones of MaSp1- and Masp2-like spider silk proteins

30

For the purposes of further classification and structure/function analyses of the novel spider silk proteins of the present invention, fibroin sequences were allocated to different ortholog groups of *Nephila*

clavipes. Silk fibroins are long proteins, comprised largely (>90%) of ensemble repeat units which are internally repetitive (4, 5, 8, 9). Ensemble repeat units from different fibroins vary in length, sometimes  
5 by an order of magnitude. It is difficult to make residue-to-residue homology statements between molecules because of this length variation and the overall modular structure of silk proteins. The gross similarities and differences between ensemble repeat units were,  
10 therefore, initially used to sort araneoid fibroins into four classes. The following proteins correspond to the MaSp1-like type of fibroin previously described in *N. clavipes* (5, 9).

MaSp1-like group proteins were characterized by  
15 short ensemble repeats with single polyalanine stretches. The remainder of a repeat was comprised of numerous GGX motifs and scattered GA motifs. The MaSp1-like group of spider silk proteins includes *N. clavipes* MaSp1, and proteins encoded by a genomic MaSp1 clone from *Argiope*  
20 *aurantia* (*A. aurantia*) (SEQ ID NO: 1), an *A. trifasciata* MaSp1 cDNA (SEQ ID NO: 2), a *Latrodectus geometricus* (*L. geometricus*) MaSp1 cDNA (SEQ ID NO: 3), a genomic MaSp1 clone from *N. madagascariensis* (SEQ ID NO: 4), a genomic MaSp1 clone from *N. senegalensis* (SEQ ID NO: 5), a  
25 genomic MaSp1 clone from *Tetragnatha kawaiensis* (*T. kawaiensis*) (SEQ ID NO: 6), and a genomic MaSp1 clone from *T. versicolor* (SEQ ID NO: 7). Amino acid sequences encoded by SEQ ID NOs: 1-7 are provided in SEQ ID NOs: 29-35, respectively.

30 MaSp2-like group proteins were characterized by short ensemble repeats comprised of one polyalanine stretch and various iterations of GPG(X)<sub>n</sub> and GP motifs. The MaSp2-like group of spider silk proteins includes *N. clavipes* MaSp2, and proteins encoded by a genomic MaSp2

clone from *A. aurantia* (SEQ ID NO: 8), an *A. trifasciata* MaSp2 cDNA (SEQ ID NO: 9), a genomic MaSp2 clone from *A. trifasciata* (SEQ ID NO: 10), a genomic MaSp2 clone from *Gasteracantha mammosa* (SEQ ID NO: 11), a *L. geometricus* MaSp2 cDNA (SEQ ID NO: 12), a genomic MaSp2 clone from *L. geometricus* (SEQ ID NO: 13), two genomic MaSp2 clones from *N. madagascariensis* (SEQ ID NOs: 14-15), and a MaSp2 genomic clone from *N. senegalensis* (SEQ ID NO: 16). Amino acid sequences encoded by SEQ ID NOs: 8-16 are provided in SEQ ID NOs: 36-44, respectively.

### EXAMPLE III

#### Clones of flagelliform-like spider silk proteins

Based on the gross similarities and differences between ensemble repeat units, the following group of proteins was classified as flag-like type fibroins, similar to those previously described in *N. clavipes* (5, 9).

Flag-like group proteins were characterized by long ensemble repeats comprised mainly of clustered GGX and GPG(X)<sub>n</sub> motifs. Each ensemble repeat had a single "spacer" region that contained amino acids atypical of araneoid silks (5). The flag-like group of spider silk proteins includes Flag from *N. clavipes* and two proteins encoded by *A. trifasciata* Flag cDNA clones (SEQ ID Nos: 17-18). Amino acid sequences encoded by SEQ ID NOs: 17-18 are provided in SEQ ID NOs: 45-46, respectively.

### EXAMPLE IV

#### Clones of spider silk proteins comprised of divergent motifs

Two of the novel spider silk proteins described herein could not be allocated readily into one of the

four classes of araneoid fibroins previously described in *N. clavipes*. This group of fibroins includes spider silk proteins comprised of divergent repetitive motifs, a feature which may reflect the diverse ecosystems of the species from which the nucleic acid sequences encoding these fibroins were derived. This category includes proteins encoded by a *Dolomedes tenebrosus* (*D. tenebrosus*) fibroin 1 cDNA (SEQ ID NO: 19) and a *D. tenebrosus* fibroin 2 cDNA (SEQ ID NO: 20). Amino acid sequences encoded by SEQ ID NOs: 19-20 are provided in SEQ ID NOs: 47-48, respectively.

#### EXAMPLE V

##### Clones of spider silk proteins comprised of atypical motifs

Seven novel spider silk proteins of the present invention comprise atypical spider silk motifs unlike those described for any previously characterized araneoid fibroin. This group of fibroins includes spider silk proteins comprised of divergent repetitive motifs, a feature which may reflect the diverse ecosystems of the species from which the nucleic acid sequences encoding these fibroins were derived. This category includes proteins encoded by an *Euagrus chisoseus* (*E. chisoseus*) fibroin 1 cDNA (SEQ ID NO: 21), a *Plectreurys tristis* (*P. tristis*) fibroin 1 cDNA (SEQ ID NO: 22), a *P. tristis* fibroin 2 cDNA (SEQ ID NO: 23), a *P. tristis* fibroin 3 cDNA (SEQ ID NO: 24), a *P. tristis* fibroin 4 cDNA (SEQ ID NO: 25), a *Phidippus audax* (*P. audax*) fibroin 1 cDNA (SEQ ID NO: 26), and a *Zorocrates* sp. fibroin 1 cDNA (SEQ ID NO: 27). Amino acid sequences encoded by SEQ ID NOs: 21-27 are provided in SEQ ID NOs: 49-55, respectively.

**EXAMPLE VI****An exemplary clone of a spider silk protein comprised of  
divergent motifs**

5           A novel spider silk protein of the present invention  
comprises atypical spider silk motifs unlike those  
described for any previously characterized araneoid  
fibroin. This fibroin is comprised of highly divergent  
repetitive motifs and is encoded by a *A. trifasciata*  
10   aciniform fibroin 1 cDNA clone (SEQ ID NO: 28). Amino  
acid sequences encoded by SEQ ID NO: 28 are provided in  
SEQ ID NO: 56.

          A consensus sequence repeat of the *A. trifasciata*  
aciniform fibroin 1 protein (SEQ ID NO: 56) comprised of  
15   approximately 200 amino acids has been identified herein.

          Amino acid sequences comprising the consensus sequence  
repeat are provided in SEQ ID NO: 57. Such a consensus  
sequence is of use in a number of applications,  
including, but not limited to: 1) the generation of  
20   degenerative nucleic acid probes capable of encoding SEQ  
ID NO: 57 which can used to screen for and identify  
nucleic acid molecules encoding novel spider silk  
proteins, 2) the generation of antibodies specific for  
portions of or all of SEQ ID NO: 57 which can be used to  
25   screen for and identify novel spider silk proteins or  
derivatives thereof, and 3) utilization as a modular unit  
in the design and production of synthetic spider silk  
proteins.

30

**EXAMPLE VII****Exemplary methods for designing synthetic spider silk  
proteins and uses thereof**

The following methods for designing synthetic spider

silk proteins are based on the amino acid composition of spider silk proteins and how repetitive regions of amino acid sequences contribute to the structural/physical properties of spider silk proteins.

5           In general, synthetic spider silk proteins can be comprised of a series of tandem exact repeats of amino acid sequence regions identified as having a spectrum of physical properties. Exact repeats would comprise regions of amino acid sequences that are duplicated  
10           precisely. Alternatively, synthetic spider silk proteins can be comprised of a series of tandem inexact repeats identified as having a spectrum of physical properties. Inexact repeats would comprise regions of amino acid sequences in which at least one amino acid sequence can  
15           be altered in the basic inexact repeat unit as long as the alteration does not change the spectrum of physical properties characteristic of the basic inexact repeat unit.

          In order to increase the tensile strength of minor  
20           ampullate silk for applications where strength and very little elasticity are needed, such as bulletproof vests, the (GA)<sub>n</sub> regions can be replaced by (A)<sub>n</sub> regions. This change would increase the tensile strength. The typical MiSp 1 protein has sixteen (GA) units. Replacing eight  
25           (GA) regions, for example, with (A) regions would increase the tensile strength from 100,000 psi to at least 400,000 psi. Moreover, if the (A)<sub>n</sub> regions were as long as the (GA)<sub>n</sub> regions the tensile strength would increase to greater than 600,000 psi.

30           To create a fiber with high tensile strength and greater elasticity than major ampullate silk, the number of (GPGXX) regions can be increased from 4-5 regions, which is the range of (GPGXX) regions typically found in naturally occurring major ampullate spider silk proteins,  
35           to a larger number of regions. For example, if the



number were increased to 10-12 (GPGXX) regions, the elasticity would increase to 50-60%. If the number were further increased to 25-30 regions, the elasticity would be near 100%. Such fibers can be used in coverings for wounds (for example, burn wounds) to facilitate easier placement and provide structural support. Such fibers can also be used for clothing and as fibers in composite materials.

The tensile strength of a very elastic flagelliform silk can be increased by replacing some of the (GPGXX) units with (A)<sub>n</sub> regions. A flagelliform silk protein contains an average of 50 (GPGXX) units per repeat. Replacing two units in each repeat with (A) regions can, therefore, increase the tensile strength of a flagelliform silk by a factor of four to achieve a tensile strength of about 400,000 psi. Uses for such flagelliform silk proteins are similar to those described for major ampullate proteins having augmented elasticity (as described hereinabove). The flagelliform proteins have additional utility in that the spacer regions therein confer the ability to attach functional molecules like antibiotics and/or growth factors (or combinations thereof) to composites comprising flagelliform proteins.

Fibers woven from combinations of the natural and/or synthetic spider silk proteins of the present invention are also encompassed herein. Such composite fibers have utility in a variety of applications, including, but not limited to, production of fabric, sutures, medical coverings, high-tech clothing, rope, and reinforced plastics.

## REFERENCES AND NOTES

1. J. Coddington, H. Levi, *Annu. Rev. Ecol. Syst.* **22**, 565 (1991).
- 5 2. F. Vollrath, *Intl. J. of Biol. Macromol.* **24**, 81 (1999).
3. J. Gosline, P. Guerette, C. Ortlepp, K. Savage, *J. Exp. Biol.* **202**, 3295 (1999).
4. P. Guerette, D. Ginzinger, B. Weber, J. Gosline,  
10 *Science* **272**, 112 (1996).
5. C. Hayashi, R. Lewis, *J. Mol. Biol.* **275**, 773 (1998).
6. P. Calvert, *Nature* **393**, 309 (1998).
7. J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, R.  
15 Lewis, *Science* **291**, 2603 (2001).
8. R. Beckwitt, S. Arcidiacono, R. Stote, *Insect Biochem. Mol. Biol.* **28**, 121 (1998).
9. Genbank# M37137, M92913, AF027735-AF027737, AF218621-AF218624.
- 20 10. K. Mita, S. Ichimura, T. James, *J. Mol. Evol.* **38**, 583 (1994).
11. Genbank# AF08334, AF095239, AF095240.
12. DNA sequences for the various fibroins were translated into amino acid sequences. For each fibroin, ensemble repeat units, higher-order aggregations of  
25 iterated sequence motifs, were aligned algorithmically then manually using MacVector (Oxford Molecular), and a consensus ensemble repeat unit was generated. The consensus ensemble repeat was defined as the most common  
30 amino acid (or gap) at each position in the alignment of repeats for each molecule. When there was a tie at an alignment position, X denoted this ambiguity in the consensus sequence.
13. P. Selden, J. Gall, *Palaeontology* **35**, 211 (1992).

14. C. Craig, *Annu. Rev. Entomol.* **42**, 231 (1997).
15. Araneoid silk sequences initially were allocated to different fibroin orthologue groups in *Nephila clavipes* (5,9) according to sequence characteristics of ensemble repeat units. These assignments were assessed through phylogenetic analyses of non-repetitive carboxy-terminal sequences, expression patterns, fit to independent phylogenetic hypotheses for Araneoidea, and minimization of gene duplication events.
16. P. Selden, *Palaeontology* **33**, 257 (1990).
17. A. Simmons, E. Ray, L. Jelinski, *Macromolecules* **27**, 5235 (1994).
18. S. Sudo et al., *Nature* **387**, 563 (1997).
19. X. Qin, K. Coyne, J. Waite, *J. Biol. Chem.* **272**, 32623 (1997).
20. A. Tatham, P. Shewry, B. Mifflin, *FEBS Lett.* **177**, 205 (1984).
21. B. Thiel, K. Guess, C. Viney, *Biopolymers* **41**, 703 (1996).
22. Q. Cao, Y. Wang, H. Bayley, *Curr. Biol.* **7**, 677 (1997).
23. J. Schultz, *Biol. Rev.* **62**, 89 (1987).
24. L. Chen, A. DeVries, C. Cheng, *Proc. Natl. Acad. Sci. USA* **94**, 3817 (1997).
25. A. Seidel, O. Liivak, L. Jelinski, *Macromolecules* **31**, 6733 (1998).
26. B. Madsen, Z. Shao, F. Vollrath, *Intl. J. Biol. Macromol.* **24**, 301 (1999).
27. G. Hormiga, N. Scharff, J. Coddington, *Syst. Biol.* **49**, 435 (2000).

While certain of the preferred embodiments of the present invention have been described and specifically

exemplified above, it is not intended that the invention  
be limited to such embodiments. Various modifications  
may be made thereto without departing from the scope and  
spirit of the present invention, as set forth in the  
5 following claims.

What is claimed is:

1. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 29 and naturally occurring allelic variants  
5 thereof.
2. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 30 and naturally occurring allelic variants  
10 thereof.
3. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 31 and naturally occurring allelic variants  
thereof.
- 15 4. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 32 and naturally occurring allelic variants  
thereof.
- 20 5. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 33 and naturally occurring allelic variants  
thereof.
- 25 6. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 34 and naturally occurring allelic variants  
thereof.
- 30 7. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 35 and naturally occurring allelic variants  
thereof.
- 35 8. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 36 and naturally occurring allelic variants  
thereof.
9. A nucleic acid molecule encoding a protein of

SEQ ID NO: 37 and naturally occurring allelic variants thereof.

5           10. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 38 and naturally occurring allelic variants  
thereof.

10           11. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 39 and naturally occurring allelic variants  
thereof.

15           12. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 40 and naturally occurring allelic variants  
thereof.

            13. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 41 and naturally occurring allelic variants  
thereof.

20           14. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 42 and naturally occurring allelic variants  
thereof.

25           15. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 43 and naturally occurring allelic variants  
thereof.

30           16. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 44 and naturally occurring allelic variants  
thereof.

35           17. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 45 and naturally occurring allelic variants  
thereof.

18. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 46 and naturally occurring allelic variants  
thereof.

5           19. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 47 and naturally occurring allelic variants  
thereof.

10           20. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 48 and naturally occurring allelic variants  
thereof.

15           21. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 50 and naturally occurring allelic variants  
thereof.

20           22. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 51 and naturally occurring allelic variants  
thereof.

23. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 52 and naturally occurring allelic variants  
thereof.

25           24. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 53 and naturally occurring allelic variants  
thereof.

30           25. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 54 and naturally occurring allelic variants  
thereof.

35           26. A nucleic acid molecule encoding a protein of  
SEQ ID NO: 55 and naturally occurring allelic variants  
thereof.

27. A nucleic acid molecule encoding a protein of SEQ ID NO: 56 and naturally occurring allelic variants thereof.

5           28. A nucleic acid molecule encoding a consensus silk protein sequence of SEQ ID NO: 56 and naturally occurring allelic variants thereof.

10           29. A nucleic acid selected from the group consisting of SEQ ID NOS: 1-28.

30. An expression vector comprising at least one of the nucleic acids of SEQ ID NOS: 1-28.

15           31. A host cell comprising the expression vector of claim 30.

32. A spider silk protein comprising SEQ ID NO: 29.

20           33. A spider silk protein comprising SEQ ID NO:30.

34. A spider silk protein comprising SEQ ID NO:31.

25           35. A spider silk protein comprising SEQ ID NO:32.

36. A spider silk protein comprising SEQ ID NO:33.

37. A spider silk protein comprising SEQ ID NO:34.

30           38. A spider silk protein comprising SEQ ID NO:35.

39. A spider silk protein comprising SEQ ID NO:36.

35           40. A spider silk protein comprising SEQ ID NO:37.



41. A spider silk protein comprising SEQ ID NO:38.
42. A spider silk protein comprising SEQ ID NO:39.
- 5 43. A spider silk protein comprising SEQ ID NO:40.
44. A spider silk protein comprising SEQ ID NO:41.
45. A spider silk protein comprising SEQ ID NO:42.
- 10 46. A spider silk protein comprising SEQ ID NO:43.
47. A spider silk protein comprising SEQ ID NO:44.
- 15 48. A spider silk protein comprising SEQ ID NO:45.
49. A spider silk protein comprising SEQ ID NO:46.
50. A spider silk protein comprising SEQ ID NO:47.
- 20 51. A spider silk protein comprising SEQ ID NO:48.
52. A spider silk protein comprising SEQ ID NO:49.
- 25 53. A spider silk protein comprising SEQ ID NO:50.
54. A spider silk protein comprising SEQ ID NO:51.
55. A spider silk protein comprising SEQ ID NO:52.
- 30 56. A spider silk protein comprising SEQ ID NO:53.
57. A spider silk protein comprising SEQ ID NO:54.
58. A spider silk protein comprising SEQ ID NO:55.
- 35

59. A spider silk protein comprising SEQ ID NO:56.

60. A spider silk amino acid consensus sequence comprising SEQ ID NO:57.

5

61. A spider silk protein selected from the group consisting of SEQ ID NOS: 29-56.

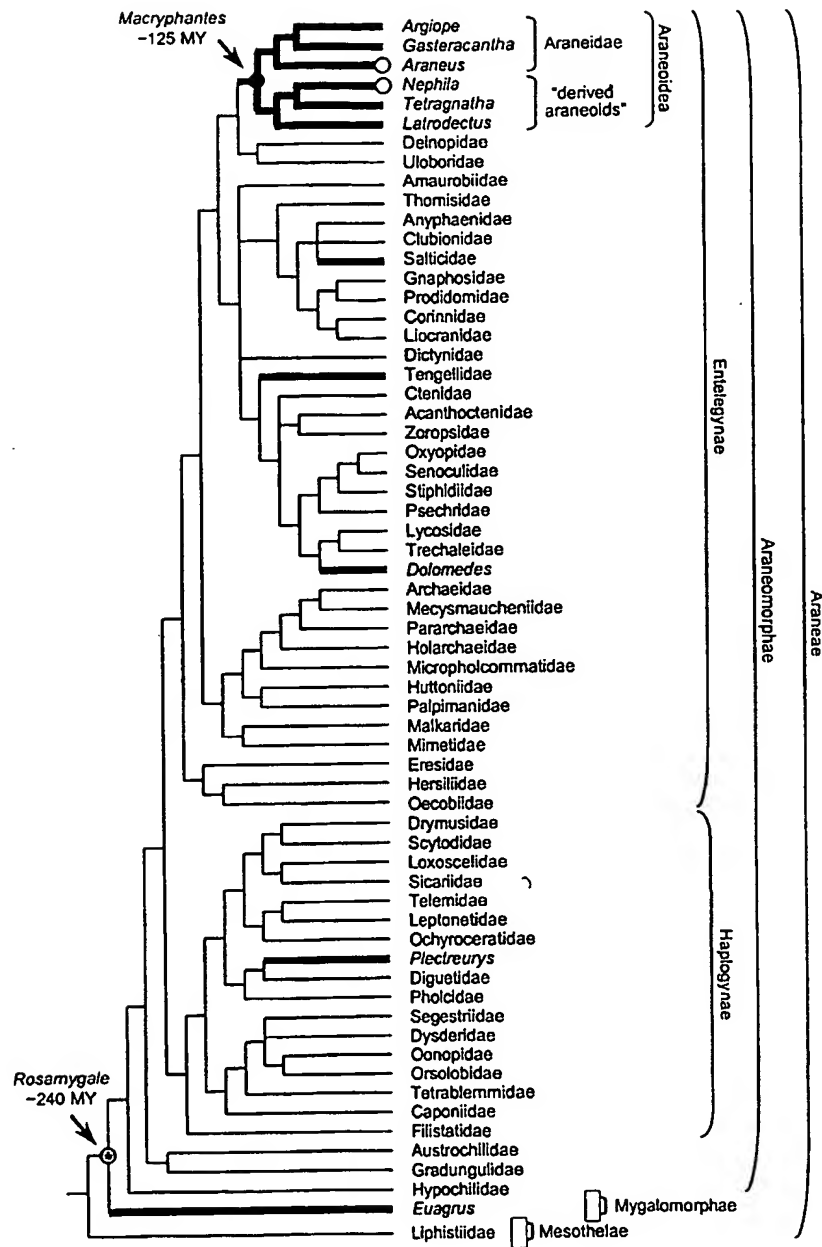
62. A silk fiber comprising at least one of the proteins of claim 61.

10

63. A copolymer fiber comprising at least two of the proteins of claim 61.

15

Figure 1





# Figure 3

MaSp1	
<i>Nep.c.*</i>	GGA--GQGGYGGLGSQGA-----GRGGLGGQ---GA--GAAAAA----
<i>Nep.m.†</i>	GGA--GQGGYGGLGSQGA-----GRGGYGGQ---GA--GAAAAA----
<i>Nep.s.†</i>	GGA--GQGGYGGLGSQGA-----GAAAAA----
<i>Tet.k.†</i>	GGLGGGQ-GAGQGGQGGAGQGGYGSGLGGAGQ-----GASAAAAAAA
<i>Tet.v.†</i>	GGLGGGQGGY-----GSLGGAGQGGQGGAGQGAASAAAAA
<i>Lat.g.<sup>Sm</sup></i>	GGA--GQGGY-----GQ-----GGQGGGA-----GAAAAA----
<i>Arg.a.†</i>	GGQ-GGXGGYGGLGSQGGAGQ-GYGSGLGGQGGAGQG-----GAAAAA----
<i>Arg.t.<sup>Sm</sup></i>	GGQ-GGQGGYGGLGXQGGAGQ-GYGAGSGGQGGXGGQ-----GAAAAA----
<i>Ara.d.*</i> (ADF-2)	GGQ-GGQGGQGGGLGSQGGAG--GAGQGGY-GAGQG-----GAAAAA----
MaSp2	
<i>Nep.c.*</i>	---GPG--QQGPGGYGPG--QQGPGGYGPGQQGPGSGPSAAAAAAA
<i>Nep.m.†</i>	---GPG--QQGPGGYGPG--QQGPGGYGPGQQGPGSGPSAAAAAAA-
<i>Nep.s.†</i>	---GPG--QQGPGXY-----GPGSGPSAAAAA----
<i>Lat.g.<sup>Sm.†</sup></i>	---GPGGYGPGPGXQQGY-----GPGSGAAAAA----
<i>Arg.a.†</i>	GGYGPAGQQGPGSGQPGSGGQQGPGGX---GPYGPSAAAAAAA-
<i>Arg.t.1<sup>Sm</sup></i>	GGYGPAGQQGPGSGQPGSGGQQGPGGQ---GPYGPSAAAAAAA-
<i>Gas.m.†</i>	GGYGPAGSGQQGPGGQPGSGGQQGPGGQ---GPYGPAAAAAAA-
<i>Ara.b.*</i>	GGYGPAGSGQQGPGGQ---GPGGQ---GPYGPASAAAAA-
<i>Ara.d.1*</i> (ADF-3)	GGYGPAGSGQQGPGGQ---GPGGQ---GPYGPASAAAAA-
<i>Nep.m.2†</i>	-GRGPGGY--GPGGQ-----GPGGPGAAAAA----
<i>Arg.t.2†</i>	-GPGGQ--GPGGQ-----GPGGYGPS--GPGASAAAAAAA-
<i>Ara.d.2*</i> (ADF-4)	-GPGGY--GPGSGGPS-----GPGAYGPG--GP-GSSAAAAAAS
MiSp	
<i>Nep.c.1*</i>	[GAGGAGGYGR--GAGAGAGAAAGAGAGAGGYGGQGGYGAGAGAGAAAAAG]- <sub>10</sub> [spacer] <sub>1</sub>
<i>Nep.c.2*</i>	[---GGYGRGVGAGAGAGAAAGAGAGAGGYGGQGGYGAGXGA---AAAGAG] <sub>10</sub> [spacer] <sub>1</sub>
<i>Ara.d.*</i> (ADF-1)	[GAGAAGGYGG--GAGAGAG-----GAGGY-GQ-GYGAGAGAGAAAAAG]- <sub>5</sub> [spacer] <sub>1</sub>
Flag	
<i>Nep.c.*</i>	[GPGGX] <sub>11</sub> [GGX] <sub>1</sub> THELDITIDGADGPITISEELTIS--GAGGS [GPGGX] <sub>11</sub>
<i>Nep.m.*</i>	[GPGGX] <sub>11</sub> [GGX] <sub>1</sub> TVIEDLITIDGADGPITISEELTIGGAGAGGS [GPGGX] <sub>11</sub>
<i>Arg.t.<sup>Sm</sup></i>	[GPGGX] <sub>11</sub> GPVTVDVDSVGGAPGG [GPGGX] <sub>11</sub> [GGX] <sub>1</sub> [GPGGX] <sub>11</sub>

# ADDENDUM I

## Nucleic Acid Sequences

The following list of standard IUPAC nucleic acid one-letter abbreviations are provided for the purposes of clarity and define the nomenclature used to indicate ambiguities in nucleic acid sequences. The nucleic acid sequences of the present invention may include at least one ambiguous nucleic acid as indicated by the inclusion of one of the below listed one-letter abbreviations.

Y	pYrimidine	(C or T)
R	puRine	(A or G)
W	"Weak"	(A or T)
S	"Strong"	(C or G)
K	"Keto"	(T or G)
M	"aMino"	(C or A)
B	not A	(C or G or T)
D	not C	(A or G or T)
H	not G	(A or C or T)
V	not T	(A or C or G)
N	aNy base	(A or C or G or T)

SEQ ID NO: 1

Argiope aurantia major ampullate spidroin 1 (MaSpl) gene, partial cds.

Genbank Accession: AF350262

DNA sequence (1344 bp)

gagccggacaaggaggagctggagccgcagctgctgcagctgcagccggtggagctgga  
ggtgctggaagaggaggattaggtgctggcggtgcaggacaaggatatggatccggatt  
aggcgggtcaaggaggagcaggtggtggcgctgccgcagctgcagcagcagcagcaggcg  
gccaaggaggacaaggtggatatggcggattaggttctcaaggtgctggtcaaggagga  
tatggagctggacaaggaggagctggagccgcagctgctgcagctgcagccggtggagc  
tggaggtgctggaagaggaggattaggtgctggcggtgcaggacaaggatacggatccg  
gattaggcgggtcaaggaggagctggtcaaggtggtgctgccgcagcagcagcagcagct  
ggcggccaaggaggacaaggtggatatggcggattaggttctcaaggtgctggtcaagg  
aggagctggtcgtggcgctgccgcagccgcagcagcagctggcggccaaggaggacgag  
gcggatatggcggattaggttctcaaggtgctggtcaaggaggatatggagctggacaa  
ggaggagctggagccgcagctgctgcagctgcagccggtggagctggagaaggaggatt  
aggtgctggcggtgcaggacaaggatatggatccggattaggcgggtcaaggaggagctg  
gtcaaggtggtgctgccgcagccgcagcagcagctggaggccaaggaggacatggtgga  
tatggcggattaggttctcaaggtgctggtcaaggaggagctggtcgtggcgctgccgc  
agccgcagcagcagctggcggtcaaggaggacaggggtggatatggcggattaggttctc  
aaggtgccgggtcaaggaggatatggagctggacaaggaggagctgcagccgcagctgct  
gcagctgcagccggtggagctggaggtgctggaagaggagaattaggtgctggcggtgc

aggacaaggatatggayccggattaggcgggtcaaggaggagctgggtcaacgtgggtgccg  
cttctgttgagcattagctggaggggcaaggaggacaaggtgggttttggcggatttagt  
tcacaaggagcaggtcaaggagcatatgggtgggtgcatacagtggacaaggagcagc  
agcatctgtttccgctgcttccgctgcagcttcacgtctgtcatcacctgggtgctgctt  
cgagagtgtcttccgctgttacttcttgggtatcaagtggcggcccaactaatctctgca  
gctttatcgaatactatcagcartgttgtttctcaaattagtgaga

SEQ ID NO: 2

Argiope trifasciata major ampullate spidroin 1 (MaSp1)

mRNA, partial cds.

Genbank Accession: AF350266

DNA sequence (1947 bp)

gcagctgcagccgcagcagcagcagccgggtggccaaggaggacaaggtggatatgacgg  
attaggttctcaaggagccgggtcaaggaggatacggacaaggaggagccgctgccgcag  
cagccgcagccagtggagctggtagtgcccaacgaggaggcttaggtgctggaggtgca  
ggacaaggatatggagccggatcaggcgggtcaaggaggagctggacaaggtggcgagc  
tgcagccacagcagcagcagccgggtggccaaggaggacaaggtggatatggcggattag  
gttccaaggatccgggtcaaggaggatacggacaaggaggagccgctgccgcagcagcc  
gcagccagtggagatgggtgggtgccggacaagaaggcttaggtgctggaggtgcaggaca  
aggatatgggtgctggattaggcgggtcaaggaggagctggacaaggtggcgagctgcag  
ccgcagcagcagcagccgggtggccaaggaggacaaggtggatatggcggattaggttct  
caaggagccgggtcaaggaggatacggacaaggaggagccgctgccgcagcagccgcagc  
cagtggagctgggtggcgccggacaaggaggcttaggtgctgcaggtgcaggacaaggat  
atgggtgccggatcaggcgggtcaaggaggagctggacaaggtggcgagctgcagctgca  
gcagcagcagccgggtggccaaggaggacaaggtggatatggcggattaggttctcaagg  
agccgggtcaaggaggatacggacaaggaggagtcgctgctgcagcagccgcagccagtgc  
gagctgggtgggtgccggacgaggaggcttaggtgctggaggtgcaggacaagaatatgg  
gccgtatcaggcgggtcaaggaggagctggacaaggtggcggaagctgcagccgcagcagc  
agcagccgggtggccaaggaggacaaggtggatatggcggattaggttctcaaggagccg  
gtcaaggaggatacggacaaggaggagccgctgccgcagcagcagcagccagtggagct  
gggtgggtgccagacgaggaggcttaggtgctggaggtgcaggacaaggatatgggtgccc  
attaggtgggtcaaggaggagcaggacaaggtagcgcatctgcagccgcagcagcagcag  
ccgggtggccaaggaggacaaggtggatatggcggattaggttctcaaggatccgggtcaa  
ggaggatacggacaaggaggagccgctgccgcagcagccgcagccagtggagctgggtgg  
tgccggacgaggaagcttaggtgctggaggtgcaggacaaggttatgggtgctggattag  
gcgggtcaaggaggagctggacaaggtggcgagctgcagccgcagcagcagccgggt  
ggccaaggaggacaaggtggatatggcggattaggttctcaaggagctgggtcaaggagg  
atacggacaaggaggagccgctgccgcagcagcatcagccgggtggccaaggagggaag  
gtggatatgggtggattaggttctcaaggagccgggtcaggaggagatatgggtgggtgggca  
ttcagtgggccaacaaggcggagcagcatctgttgccactgcttccgctgctgcttcacg  
cttgtcatcacctgggtgctgcttcgagagtttcttctgccgttacatctttgggtgtcaa  
gtgggtggcccaactaattctgcagctttatctaatactatcagcaatgttgtttcacia  
aattgtttcagctttggtacatatcttgggtcagctaacattggacaagttaactcca  
gcgggtgttgggcgatcagcttctattgtgggacaatctataaaccagctttctcataa

SEQ ID NO: 3

*Latrodectus geometricus* major ampullate spidroin 1 (MaSp1)  
mRNA, partial cds.

Genbank Accession: AF350273

DNA sequence (1083bp)

```
gctggctcaggacaaggtggttatggacaaggatatggtgaaggtggtgctggacaaggg
ggagcaggagccgcagcagcagccgctgcagcagctggtggagctggacaaggtggacaa
ggcggttatggacaaggatatggtcaaggtggtgccggacaaggtggagcaggagccgca
gcagcagctgcagctggtggagctggacaaggaggctacggccgaggtggagcaggacaa
ggtgcagcagcggcagcagcagctgcaggttcaggacaaggaggacaaggtggttatgga
caaggttatggtcaaggtggtgctggacaagggggagcaggagccgcagcagcagcagct
gcagctggtggagctggacaaggaggatacggacgaggaggagcaggacagggaggagca
gctgcagccgctgcagcagctggaggagccggtcaaggtggacaaggcggttatggacaa
ggatatggtcaaggtggtgccggacaggggtggagcaggagccgcagcagcagcagctgca
gctggtggagctggacaaggaggatatggccgaggtggagcaggacaaggtggatcagca
gcagccgcagcagcagctggtggagcaggacaaggaggatatggccgaggtggtgccgga
caaggtggagcaggttcagcagcggcagcagctgcagctggcggttctggacaaggagga
caaggtggttatggacaaggatatggtcaaggtggtgctggacaaggtggagcagctgct
gcagcttctgctttggcagctccagctacaagtgcgagaatttcttctcatgcctcgact
cttctttcaaattggtcccaaatccagcttcaatttcaaattgtattagtaatgctgta
tcccaaattagttcgagcaatccaggagcttcttcgtgtgacgttcttgttcaagctctt
cttgaacttgtcacagcgttactcaccattattgggtccttcaatgttggcaatgttaat
tatgattcttcaggccaatatgcacaagtggtttcacagtccgtgcaaaacgcatttgtt
taa
```

SEQ ID NO: 4

*Nephila madagascariensis* major ampullate spidroin 1 (MaSp1)  
gene, partial cds.

Genbank Accession: AF350277

DNA sequence (703 bp)

```
gaggtcttgggtggacaaggtgcaggacaaggagctggagcagcagcagcagcagctggtg
gtgccggacaaggaggatatggaggtcttggagccaaggtgctggccgagggcgatatg
gtggacaaggagctggagcagcagcagccgctgccgcaggaggtgccggacaaggaggat
atggaggtcttggagccaaggtgctggacaaggaggatacggaggtcttgggtggacaag
gtgcaggacaaggagcagcagcagcagcagcagctggtggtgccggacaaggaggatatg
gaggtcttggagccaaggtgctggccgagggcgatatggtggacaaggtgcaggagcag
cagcagctgcaactggtggtgctggacaaggaggatatggtggtgctcggttctggggcgt
ctgctgcctctgcagctgcatcacgtttgtcttctcctcaagctagtccaagagtttcat
cagctgtttccaacttgggttgcaagtggctctacgaattctgcggcattgtcaagtacaa
tcagtaacgcggtttcacaattggcgccagcaatcctggtcttcttggtatgtgatgtcc
tcattcaagctcttctcgaggttgtttctgctcttatccatatcttaggttcttccagca
tcggccaagttaattatggttccgctggtcaagccactcagat
```



SEQ ID NO: 5

*Nephila senegalensis* major ampullate spidroin 1 (MaSp1)  
gene, partial cds.

Genbank Accession: AF350279

DNA sequence (763 bp)

gaggtcttggtggacaaggtgctggacgaggagctggagcagccgctgcagcagctggag  
gtgctggacaaggaggatcacggaggtcttggtggacaaggagctggagccgctgccgcag  
cagcgggtggtgccggacaaggaggacaaggattaggtggaagaggtgcagcagcagctg  
gaggtgctggacaaggaggatcacggaggtcttggtggacaaggtgctggacgaggagctg  
gagcagccgctgcagcagctggaggtgctgggtcaaggaggatcacggaggtcttggtggac  
aaggagctggagcagcagcagcagccgctgcagcaggaggtgctggacaaggagggatg  
gaggtcttggaagccaaggtgctggacgaggaggatattggaggacaaggtgcaggagcgg  
cagtagcagcgattggtggcggttgacaaggaggctatggtggtgctcggttctggggcgt  
ctgctgcctctgcagctgcttctcgcttcttctcccgaagctagttcaagagtttcat  
ctgctgtttccaacttggtttcaagtggtcctactaattctgcggcattgtcaagtacta  
tcagtaatgtggtctcacaaataggcgccagcaatcctggtcttcttggtatgtgatgtcc  
tcattcaagctcttctcgaaagttgttctgctcttgtccatatcttaggctcttcagca  
tcggccaggttaattatggttccgctggtcaagccactcagat

SEQ ID NO: 6

*Tetragnatha kauaiensis* major ampullate spidroin 1 (MaSp1)  
gene, partial cds.

Genbank Accession: AF350285

DNA sequence (853 bp)

gatccggactcggaggagcaggacaaggagccggccaaggagcatcagctgccgccgcmg  
cagcagcagsaggaggccttgagggtggccaaggagcaggtcaaggaggacaacaaggtg  
cyggacaaggaggctacggatccggactcggaggagcaggacaaggagcatcagctgccg  
ccgcagcagcagcagcaggaggccttgagggtggccaaggagcaggtcaaggaggacaac  
aaggtgctggacaaggaggctacggatccggactcggaggagcaggacaaggagcatcag  
ctgccgccgcagcagcagcagcaggaggccttgagggtggccaaggagcaggtcaaggag  
gacaacaaggtgctggacaaggaggctacggatccggactcggaggagcaggacaaggag  
ccggccaaggagcatcagctgccgccgcagcagcagcaggaggccttgagggtggccaag  
gaggttatggttctggtcttgagggtgtaggacaaggagggaagggtttaggtgggt  
caagaaactccgcaactaatgcaatttcttaattctgcctctaacgctgtctcacttctct  
catcacctgcttcaaagtcaagaatttcttctgctgtgtctgccttggcatccggtgcag  
catctggtcctggatatttatctagcgttatcagtaatgttggttctcaagtcagctcaa  
acagtgggtggacttggttggtgcgatactcttggtcaagctcttcttgaagctgctgctg  
ctcttggtgcatgtattggcttcttctagtggtggacaagtcaaccttaacacagcgggat  
acacttctcaact

SEQ ID NO: 7

*Tetragnatha versicolor* major ampullate spidroin 1 (MaSp1)  
gene, partial cds.

Genbank Accession: AF350286

DNA sequence (535 bp)

```
gatctggacaaggagcatccgccgctgcggcagcagcaggaggccttggaggtggacaag
gaggttacggttctggtctaggaggtgcaggacaaggaggacaacaaggagctggacaag
gagcagcagctgccgcagcatcagcagcagcaggaggccttggaggtggacaaggaggtc
aacaaggagcaggccgaggtggactacaaggagctggacaaggaggacaagggtgctctag
gtggatcaagaaactccgcagctaattgcagtttcacgtctctcttcacctgcttcaaatg
caagaatttcttctgctgtgtctgccttggcatccggtggagcatctagtcgccgatatt
tatctagcattattagcaatgtggtttctcagggttagctcaaacaatgatggactttctg
ggtgcgacactgttgttcaagctcttcttgaagttgctgctgctcttgtgcatgtattgg
cttcttctaataattgggcaagtcaaccttaatactgccggatacacttcccaact
```

SEQ ID NO: 8

*Argiope aurantia* major ampullate spidroin 2 (MaSp2) gene,  
partial cds.

Genbank Accession: AF350263

DNA sequence (1049 bp)

```
accaggmggtgccggccaacaaggtcctggcgggtcaaggaccatacggaccagggtgcag
ccgccgcagcagcagccgctggaggatatggaccaggagctggacaacaaggcccagrt
ggagccggacaacaaggacccggtcccaaggaccaggaggtgccggtcaacaaggacc
tggtggacaaggaccatacggaccaggagcagccgccgcagcagcagcagtaggaggat
ayggaccaggagctggacaacaaggacctggaagtcaaggaccaggaagtgggtggaca
caaggacctgggtgggtcaaggaccttatggaccaagtgcagccgccgcagcagccgc
tggaggctatggaccaggagctggacaacaaggacctggaagtcaaggaccaggaagt
gtggacaacaaggacctgggtggtctaggaccttatggaccaagtgcagccgcagcagca
gcagccgctggaggctatggaccaggagctggacaacaaggacctggaagtcaaggacc
aggaagtgggtggacaacaagacctgggtggtctaggaccttatggaccaagtgcagccg
cagcagcagcagccgctggaggctatggaccaggagctggacaacaaggacctggaagt
caaggaccaggaagtgggtggacaacaagacctgggtggtctaggaccttatggaccaag
tgcagccgcagcagcagcagccgctggaggctatggaccaggagctggacaacaaggac
ctggaagtcaaggaccagttgcatccgcagcagcctctcgtctttcttctcctcaaggc
agttctagagtttcatctgctgtgtcaactttgggtgtcgagtgggtcctacgaatcctgc
cgcactttctaattgctatcagtagcgttgatcacaaagttagtgcaagtaatcctggtc
tttctgggtgtgacgttctcgttcaagcattgctggaacttgatccgcccttgtagac
atccttgggtcttccagcattgggcaaatattacgccgcgtctt
```

SEQ ID NO: 9

*Argiope trifasciata* major ampullate spidroin 2 (MaSp2)

mRNA, partial cds.

Genbank Accession: AF350267

DNA sequence (1336 bp)

```
cgctggaccaggatacggaccaggagccggacaacaaggacctggaagtcaaggaccag
gaagtgggtggacaacaaggacctgggtggacaaggaccatatggaccaagcgctgccgcc
gcagcagctgccgctggaccaggatatggaccaggagctggacaacaaggaccagggaag
tggcggacaacaaggaggccaaggatctggacagcaaggaccaggaggtgccggtcaag
gaggtcctcgtgggtcaaggaccatacggaccagggtgcagccgccgccgcagctgct
ggaggatacggaccaggagctggacaacaaggacctggaagtcaaggaccgggaagtgg
tggacaacaaggctcctggtagtcaaggaccatatggaccaagtgcagccgcagcagcag
cagccgctggaccaggatacggaccaggagccggacaacaaggacctggaagtcaagga
ccaggaagtgggtggacaacaaggacctgggtggacaaggaccatatggaccaagcgatgc
cgccgcagcagctgccgctggaccaggatatggaccaggagctggacaacaaggaccag
gaagtggcggacaacaaggaggccaaggatctggacagcaaggaccaggaggtgccggt
caaggaggtcctcgtgggtcaaggaccatacggaccagggtgcagccgccgccgcagc
tgctggaggatacggaccaggagctggacaacaaggacctggaagtcaaggaccggaa
gtggtggacaacaaggctcctggtagtcaaggaccatatgggccaagtgcagccgcagca
gcagcagccgctggaccaggatacggaccaggagccggacaacaaggacctggaagtca
aggaccaggaagtgggtggacaacaaggctcctggtagtcaaggaccatatggaccaagtgcagccgcagcagcagcagccgctggaccaggatacggaccaggagccggacaacaaggac
cctggaagtcaagcaccagttgcacccgagctgcttctcgactttcttctcctcaagc
cagttctcgagtttcatcagctgtgtcaactttgggtgtcgagcggtcctacgaatcctg
cctcactctctaatactatcagtagcgttgatcacaagtcagttcaagtaatcctggt
ctttctggttgcatgtactcgtccaagcattgctggaaattgtatccgccctggtaca
tatccttggtatcttctagcattgggcaaattaattacgccgcttcttctcagtatgcgc
aattggttggtcaatctttaactcaagcccttggttga
```

SEQ ID NO: 10

*Argiope trifasciata* major ampullate spidroin 2-like protein  
gene, partial cds.

Genbank Accession: AF350268

DNA sequence (360 bp)

```
ggacctggacaacaaggacctggagggtatggaacatccggacctggaggtgcttctgc
cgccgccgctgctgcagctgcaggtggacctggaggacaaggaccatctggaccaggac
caccaggaccaggaggatatggaccatccggaccaggagcagccgcagccgccgctgca
gcagcaggtggacccggaagtcaaggacctggacaacaaggacccggaggctacggacc
atctggacctggaggagcttctgccgccgccgctgctgcagctgcaggtggacccggag
gtcaaggatcatacggaccaggacaacaaggaccagggtgcaggacaatacggacccgga
caacag
```

SEQ ID NO: 11

*Gasteracantha mammosa* major ampullate spidroin 2 (MaSp2)  
gene, partial cds. (note: some consider *G. mammosa* a  
subspecies of *G. cancriformis*)  
Genbank Accession: AF350272  
DNA sequence (1026 bp)

```
ggccaacaaggacctggaagtcaaggaccatacggacctggtgcagcagctgccgcagc
agcagcagctggaggataccgacctgtatctggtcaacaaggacctggacaacaaggac
caggaagcgggtggccaacaaggacctggaggccaacgaccttacggaccaggtgcagcc
gcagcagcagcagccgcaggaggatacggacctggatctggacaaggaggacctggaca
acaaggaccaggaagtggcggacaacaaggacctggagggtcaaggaccatacggacctg
gtgcagccgccgcagcagcagcagccgcaggcggatacggacctggatctggacaagga
ggacaacaaggacctggatcacaaggaccaggaagtgggtggacaacaaggacctggggg
acaagggtccatacggacctagtgcgctgcagcagcagcagccgttgaggatacggac
caggagctggacagcaaggacctggacaacaaggaccaggaagtgggtggccaacgagga
cctggagggtcaaggaccatatggaccaggagcagcagctgccgcagcagcagcagctgg
tggatatggacctgcatctggtcaacaaggacctggacaacaaggaccaggaagtgggtg
gccaacgaggacctggagggtcaaggaccatatggaccagggtgcagcagcagcagcagctc
gcaggaggatatggaccaggaagtgggtggaagccctgcatcaggagcagcttctcgact
ttcttctcctcaagccggtgccagagtttcttcagctgtatcagcccttgtcgcaagtg
gccccactagtccagctgctgtttccagcgccatcagtaatggtgcatcacaattagtg
gcaagcaatcctgggtctttccggctgcgatgttcttgtacaagcattacttgagattgt
atcagctcttgatctattctctcatccgctagtatcggacaaatcaattatggcgcat
ctggtcaatatgccgccatgatt
```

SEQ ID NO: 12

*Latrodectus geometricus* major ampullate spidroin 2 (MaSp2)  
mRNA, partial cds.  
Genbank Accession: AF350274  
DNA sequence (1118bp)

```
gcatctgcgtctggtggagcaggacctggaagacaacaaggatatggaccaggaggatca
ggagcctcggcagcagcagccgccgctggaggagctggcccaggaggatacggacaa
ggaccatctggttacggcccatctggacctggtgcacaacaagggttacggaccaggaggc
caaggaggatctggagcagcagctgcagcagccgcagcagcaggtcttgacctggagga
tatggaccaggagcagcaggaccaggaagttaggtccaagtggacctggaggatctggt
gcagctgccgcagccgctgctgctagtggaccaggaggacaacaaggatatggaccagga
ggaccaggagcctcagcagcagcagccgccgctggaggatctggacctggaggatac
ggacaaggaccatctggttacggcccatctggacctggtgcacaacaagggttacggacca
ggaggccaaggaggatctggagcagcagctgcagcagccgcagcagcaggtcttgacgt
ggaggatatggaccaggagcagcaggaccaggaattatggtccaagtggacctggagga
tctggtgcagctgcctcagccgctgctgctagtggaccaggaggacaacaaggatacggg
ccagggtggatctggagcagctgctgcagccgctctggtggagcaggacctggaagaca
caaggatatggaccaggaggatcaggagccgcagcagcagcagccgccgctggagga
tctggtccaggaggatacggacaaggaccagccggttacggaccaggaggccaaggagga
tccggaggagcagctgcagcagccgcagcagcaagctctggacccggaggatatggacca
```

ggagcagcaggaccaggcaattatggtccaagtggacctgggggatctggtgcagctgcg  
gcagctgctgctgctagtggaccaggaggacaacaaggatacggaccaggtggatctggt  
gcattcgcagcagcagcggctggtggtgcaggacctggaagacaacaagcataatggacct  
ggaggatcaggagctgcagcagcagcagcagtgatggatc

SEQ ID NO: 13

Latrodectus geometricus major ampullate spidroin 2 (MaSp2)  
gene, partial cds.

Genbank Accession: AF350275

DNA sequence (1196 bp)

gcgagaccaggaagttatggtccaagtggacctggaggatctggtgcagctgccgcagcc  
gctgctgctagtgaccaggaggacaacaaggatatggaccaggaggaccaggagcctca  
gcagcagcagccgccgccgctggaggatctggacctggaggatacggacaaggaccatct  
ggttacggcccatctggwcttggtgcacaacaagggttacggaccaggaggccaaggagg  
tctggagcagcagctgcagcagccgcagcagcaggtcttgacctggaggatatggacca  
ggagcagcaggaccaggaaattatggtccaagtggacctggaggatctggtgcagctgcc  
tcagccgctgctgctagtgaccwggaggacaacaaggatacggaccagggtgatctgga  
gcagctgctgcagccgcgtctggtggagcaggacctggaagacaacaaggatatggacca  
ggaggatcaggagccgcagcagcagcagccgccgctggaggatctggtccaggaggag  
tacggacaaggaccarscgttacggaccaggaggccaaggaggatccggaggagcagct  
gcagcagccgcagcagcaagctctggacccgraggatatggaccaggagcagcaggacca  
ggmaattatggtccaagtggacctggrggatctggtgcagctgcggcagctgctgctgct  
agtggaccaggaggacaacaaggatacggaccagggtgatctggtgcatctgcagcagca  
gcggctggtggtgcaggacytgaagacaacaagcatatggacctggaggatcaggagct  
gcagcagcagcagcagtgatcgggagggttacggycccgcaatatggtgccagctcc  
gttgcttctagcgtgctgctgcggcctcggcattatcttctctaccacgcattgctaga  
atttcttcccatgctcaactttattatcaagtggaccaactaactctgcagctatttct  
aatgtcattagcaatgctgtttccaagtcagtgcaagcaatccaggatcttctctctgt  
gatgtccttggttcaagcaacttcttgaattgattactgcmttaattagcatagtggattct  
tctaacattggacaagttaattacggttcttcaggccagtatgcgcaaatggttgg

SEQ ID NO: 14

Nephila madagascariensis major ampullate spidroin 2-like protein gene, partial cds.

Genbank Accession: AF350276

DNA sequence (5858 bp)

ggaggttatggacaaggaccatcaggatatgctcaaggatcatctgctgccagtgcagcg  
gcacctagtggatacgtcccaagccaaacaggccagctctggactgggagcagcagcagca  
gcagctgctgttgccctagtgggtacggcccaagtcaacaaggaccatctggaccagga  
gctgctacagccgccgagctggaagaggaccogaagggttacggaccagacaacaaggga  
cctggtgcaacagcagccgcagctggacctggagggttacggaccagacaacaaggacct  
ggaggctacggacctggacaacaaggacctggtgcagctgcagccgccgctgcaggacga  
ggacctggagggttacggaccgcgacagcaaggaccaggaggaccgggtgcagcagcagcc  
gcagctggatctgaaggctacggtcccggaacaagaaggccaagaggacctggtgcagcc  
gcagctggacctggaggctacggacctggacaacaaggagctagtgcagctgcacccgc

gctgcaggacgaggacctggaggctacggtcccggacaacaaggaccaggaggacctagt  
gcagccgcagctggacctggaggctacggacccggacaacaaggaccaagtgcagctgca  
gccgccgctgctggaagtggctcctggagggttacggacccggacaacaagggtccaggagga  
cccgggtgcagcagcagccgcggctggacctggagggttacggacctggacaacaaggacct  
ggtgctgccgcagcagcagcaggacgaggacctggagggttacggacctggtcaacaagga  
ccaggaggacctggtgcagccgccgcagcagcagcaggaagaggaccaggagggttacgga  
cccggacaacaaggaccaggaggacccgggtgcagcagcagccgccgctggaccaggagga  
tacggacctggagggttacggacccggacaacaaggaccaggaggacctggtgcagccgcc  
gcagcagcagcaggaagaggaccaggagggttacggaccaggacaacaaggaccaggaca  
caaggaccagggggatctggtgcagcagcagccgctgcaggacgagggcctggagggttac  
ggacccggacaacaaggaccaggaggacccgggtgcagcagcagccgccgctggaccagga  
ggatacggacctggacaacaaggacctggtgccgccgccgcagcagcagcagcaggaagga  
ggacctggagggttacggacccggacaacaaggaccaggaggacctggtgcagccgccga  
gcagcagcaggaagaggaccaggagggttacggaccaggacaacaaggaccaggacaaca  
ggaccaggaggatctggtgcagcagcagccgctgcaggacgaggacctggagggttacgga  
cccggacaacaaggaccaggaggacccgggtgcagcagcagccgccgctggaccaggagga  
tacggacctggacaacaaggacctggtgcagccgccgcagcagcagcagcaggaagagga  
ccaggagggttacggaccaggacaacaaggaccaggaggatctggtgcagcagcagccgct  
gcaggacgaggacctggagggttacggaccaggacaacaaggaccaggaggacccgggtgca  
gcagcagcagccgctggacctggagggttacggacctggacaacaaggaaactggtgcagct  
gcagccgccgctgctggaagtgggtgccggagggttatggaccggacaacaaggaccagga  
ggacctggtgcagcagcagccgcagctggacctggaggatacggacctggacaacaagga  
cctggtgcagctgcagccgccgctgctggaagtgggtcccggagggttatggaccggaca  
caaggaccaggaggatccagtgcagcagcagccgccgctggaccaggacgatacggacct  
ggacaacaaggacctggtgcagctgcagccgctctgctggaagaggaccaggagggttac  
ggacccggacaacaaggaccaggaggacctggtgcagcagcagccgcagctggacctgga  
ggatacggacctggacaacaaggacctggtgcagctgcagccgccgctgctggaagtgggt  
cccggagggttatggacctggacaacaaggaccaggaggacctggtgccgccgccgcagca  
gcagcaggaagaggaccaggagggttacggacaaggacaacaaggaccaggaggacctggt  
gcagcagcagccgcagctggacctggaggatacggacctggacaacaaggacctggagga  
gctgcagccgccgctgctggaagtgggtcccggagggttatggacccggacaacaaggacca  
ggaagatctggtgccgccgccgcagcagcagcagcaggaagaggaccaggagggttacgga  
cccggacaacaaggaccaggaggacccgggtgcagcagcagccgccgctggaccaggagga  
tacggacctggacaacaaggacctggtgccgccgccgcagcatcagcaggaagaggacca  
ggagggttacggaccaggacaacaaggaccaggaggatctggtgcagcagcagccgctgca  
ggacgaggacctggagggttacggacccggacaacaaggaccaggaggacctggtgcagcc  
gccgcagcagcagcaggaagaggaccaggagggttacggaccaggacaacaaggaccagga  
caacaaggaccaggaggatctggtgcagcagcagccgctgcaggacgaggacctggagggt  
tacggacccggacaacaaggaccaggaggacccgggtgccgcagcagccgccgctggacca  
ggaggatacggacctggacaacaaggacctggtgcagctgcagccgccgctgctggaagt  
ggtcccggagggttacggacccggacaacaaggaccaggaggacctggtgcagcagcagcc  
gctgcaggacgaggacctggagggttacggacctggtcaacaaggaccaggaggacctggt  
gcagccgccgcagcagcagcaggaagaggaccaggagggttacggaccaggacaacaagga  
ccaggacaacaaggaccagggggatctggtgcagcagcagccgctgcaggacgagggcct  
ggagggttacggacccggacaacaaggaccaggaggacccgggtgcagcagcagccgccgct  
ggaccaggaggatacggacctggacaacaaggacctggtgccgccgccgcagcagcagca  
gcaggacgaggacctggagggttacggacccggacaacaaggaccaggaggacctggtgca

gccgccgcagcagcagcagcaggaagaggaccaggagggttacggaccaggacaacaaggacca  
ggacaacaaggaccaggaggaccgggtgcagcagcagcagccgctggacctggagggttac  
ggacctggacaacaaggaaactgggtgcagctgcagccgccgctgctggaagtgggtgccgga  
ggttatggacccggacaacaaggaccaggaggacctgggtgcagcagcagccgcagctgga  
cctggaggatacggacctggacaacaaggacctgggtgcagctgcagccgccgctgctgga  
agtgggtcccgagggttatggacccggacaacaaggaccaggaggatccagtgcagcagca  
gccgccgctggaccaggacgatacggacctggacaacaaggacctgggtgcagctgcagcc  
gccgctgctggaagtgggtcccgagggttatggacccggacaacaaggaccaggaggacct  
gggtgccgccgccgcagcagcagcagcaggaagaggaccaggagggttacggaccgggaca  
caaggaccaggaggacctgggtgcagcagcagccgcagctggacctggaggatacggacct  
ggacaacaaggacctgggtgcagctgcagccgccgctgctggaagtgggtcccgagggttat  
ggacctggacaacaaggaccaggaggacctgggtgccgccgccgcagcagcagcaggaaga  
ggaccaggagggttacggacaaggacaacaaggaccaggaggacctgggtgcagcagcagcc  
gcagctggacctggaggatacggacctggacaacaaggacctggagcagctgcagccgcc  
gctgctggaagtgggtcccgagggttatggacccggacaacaaggaccaggagaatctggt  
gccgccgccgcagcagcagcagcaggaagaggaccaggagggttacggaccgggacaaca  
ggaccaggaggaccgggtgcagcagcagccgccgctggaccaggaggatacggacctgga  
caacaaggacctgggtgccgccgccgcagcatcagcaggaagaggaccaggagggttacgga  
ccaggacaacaaggaccaggaggatctggtgcagcagcagccgctgcaggacgaggacct  
ggagggttacggaccgggacaacaaggaccaggaggacctgggtgcagccgccgcagcagca  
gcaggaacaggaccaggagggttacggaccaggacaacaaggaccaggaggatctggtgca  
gcagcagccgctgcaggacgaggacctggagggttacggaccgggacaacaaggaccaggga  
ggaccgggtgccgcagcagccgccgctggaccaggaggatacggacctggacaacaaggga  
cctgggtgcagctgcagccgccgctgctggaagtgggtcccgagggttacggaccgggaca  
caaggaccaggaggacctgggtgcagcagcagccgctgcaggacgaggacctggagggttac  
ggacctggtcaacaaggaccaggaggacctgggtgcagccgccgcagcagcaggaaga  
ggaccaggagggttacggaccaggacaacaaggaccagggggatctggtgcagcagcagcg  
gcttcaggacgagggcctggagggttacggaccgggacaacaaggaccaggaggaccgggt  
gcagcagcagccgccgctggaccaggaggatacggacctggacaacaaggacctgggtgcc  
gccgccgccgcagcagcagcaggaacgaggacctggagggttacggaccgggacaacaaggga  
ccaggaggacctgggtgcagccgccgcagcagcagcaggaagaggaccaggagggttacgga  
ccaggacaacaaggaccaggaggatctggtgcagcagcagccgctgcaggacgaggacct  
ggagggttacggaccgggacaacaaggaccaggaggacctgggtgcagcagcagccgccgct  
ggaccaggaggatacggacctggacaacaaggacctgggtgcagccgcagcagcagcagga  
agaggaccaggagggttacggaccaggacaacaaggaccaggaggatctggtgcagcagca  
gccgctgcaggacgaggacctggagggttacggaccgggacaacaaggaccaggaggacc  
ggtgcagcagcagccgccgctggacctggagggttacggacctggacaacaaggaaactggt  
gcagctgcagccgccgctgctggaagtgggtcccgagggttatggacccggacaacaaggga  
ccaggaggacctgggtgcagcagcagccgccgctggacctggaggatacggacctggaca  
caaggacctgggtgcagctgcagccgccgctgctggaagtgggtcccgagggttatggacc  
ggacaacaaggaccaggaggacctgggtgcagccgccgcagcagcagcaggaagaggacca  
ggagggttacggaccaggacaacaaggaccaggaggatc

SEQ ID NO: 15

*Nephila madagascariensis* major ampullate spidroin 2 (MaSp2)  
gene, partial cds.

Genbank Accession: AF350278

DNA sequence (1692 bp)

```
aacagggaccatctggacctggaagtgcagcggcagcggcagcagcaggacctggacaac
aaggaccaggaggatatggaccaggacaacaaggtccaggaggatacgggtccaggacaac
aaggaccatctggaccaggcagtgagctgcagcagcagcagccgccgagcaggacctg
gacaacaaggaccaggaggatatggaccaggaccacaaggcccaggaggatatggaccag
gacaacaaggtccatcaggatatggaccaggacaacaaggtccatctggaccaggcagtg
cagcttcagcagccgcagcagcaggatctggacaacaaggaccaggaggatatggaccag
gacaacaaggaccaggaggatatggaccaggacaacaaggaccatctggaccaggtagtgc
cggcagcagcagccgcagcaggaccaggacaacaaggcccaggaggatatgggtccaggac
aacaaggtccaggaggatatggaccaggacaacaaggaccatctggaccaggtagtgcag
ctgcagcagccgccgcagcaggacctggacaacaaggaccaggaggatatggaccaggac
aacaaggtccaggaggatatggaccaggacaacaaggaccatctggacctggaagtgcag
cggcagcggcagcagcaggacctggacaacaaggaccaggaggatatggaccaggacaac
aaggtccaggaggatatgggtccaggacaacaaggaccatctggaccaggcagtgagctgc
cagcagcagccgccgcagcaggacctggacaacaaggaccaggaggatatgggtccaggac
aacaaggaccaggacaacaaggaccatctggaccaggtagtgcagcagcagcagccgcag
caggaccaggaccacaaggcccaggaggatatggaccaggacaacaaggcccaggaggat
atggaccatctggaccaggtagtgcagctgcagcagccgccgcagcaggacctggacaac
aaggaccaggaggatatggaccaggacaacaacgtccatcaggatatggaccaggacaac
aaggtccatctggaccaggcagtgagctgcagctgccgctgcagcagcaggacctggacaacaag
gaccaggtgcttacggaccatcaggacctggaagtgcagcagccgcagcaggacctggag
gatatggaccagcacaacaaggaccatctggagcaggcagtgagcagctgcagctgcag
caggacctggtggatatggaccagtgcaacagggaccatctgggtcctggaagcgcagccg
gacctggaggttatggaccagcgcaacaaggaccagctcgatatggacctggaagcgcgg
ccgcagctgctgccgctgcaggatctgcaggttatgggtccaggtcctcaagcatccgctg
cagcttctcgacttgcttctccagattcaggcgctagagttgcatctgctgtttctaact
tggtatccagtggtccaactagctctgctgccttatcaagcgtcatcagtaacgctgtgt
ctcaaattggagccagtaatcctggtctctctggttgcatgtcctcattcaagctctct
tggaatcgtttcggcttggtgtaaccattctttcttcattcattggtcaagttaatt
atggagcggctt
```

SEQ ID NO: 16

*Nephila senegalensis* major ampullate spidroin 2 (MaSp2)  
gene, partial cds.

Genbank Accession: AF350280

DNA sequence (693 bp)

```
aacagggaccaggaggatatggaccatctggaccrggaagcgctgcagcagcttcagccg
cagcaggacctggacaacaaggaccaggtgcttacggaccatcaggacctggaagtgcag
cagccgcagcgggacctggagkacaggaccaggacaacaaggtccatctggaccaggag
ctgccgccgcgcagcaggacctggacaacaaggtccaggaggatacggaccaggagctg
ccgccgcygcagcagccgcagcaggacctggacaacaaggaccagttgcatacggaccat
```



caggacccggaagtgcagcctctgcagctggacctggagggttatggaccagctcgatatg  
gacctcgggaagtgcagcagcagcagccgctgctggcagcagctctgcaggttatgggc  
caggtcctcaggcatccgctgcagcttctcgtcttgcttctccagactcaggtgctagrg  
ttgcatctgctgtttctaacttggatccagtggtccaactagctctgctgccttatcaa  
gtgttatcartaacgctgtgtctcaaattggcgcaagtaatcctggctctctggttgcg  
atgtcctcattcragctctcttgaaatcggttctgcttggtgaaccatcctttcttcat  
ctagcattggtcaagttaattatggagcggctt

SEQ ID NO: 17

*Argiope trifasciata* flagelliform silk protein (Flag) mRNA,  
partial cds. (with C-term.)

Genbank Accession: AF350264

DNA sequence (1958 bp)

gtgcaggtggaccaggagcaggtggagcaggagctggtggtgctggacctggaggattt  
ggagggtccaggtggattcggaggagcggcggtcctggaggaccaggcggcccaggagg  
agcaggcgggtggtgccggcggtcgtggcggtattgtacggacctggagggtgctggaggat  
tgtacgggtcctggaggattatacggacctggaggagctggagttcccggagcgccagga  
gcttctggttagagcaggaggtatcggaggtgcagctggaggagcaggagccggtggtgt  
cggacctggaggagctctctggaggcgctggcggtgctggcggtatcaggtgtaacagttg  
tagagtcagttagtggtggagccggaggaccaggagctggtggtgctgggtcctgga  
ggtgctggacctggaggagttggaccgggaggtatttacggaccaggaggagctggagg  
actttatggaccgggtgcaggtggagccttcggaccaggaggaggagctggtgcaccag  
gaggacctggagggtccaggtggaccaggcggcccaggtggtcctggaggaggagtaggc  
ggagcaggaaccggcggtggtgttgcccaggagctggagggttgtagacctctggagg  
tgcaggtggaaccgggtccggtatctgtctcttcaactgtaagtgtcggtggtgctggcg  
gaccaggtgcaggtggaccaggagcaggtggagcaggagctggtggtgctggacctgga  
ggatttgagggtccaggtggattcgggtggagcggcggtcctggaggaccaggcggccc  
aggaggagcaggcgggtggtgccggcggtcgtggcggtattgtacggacctggagggtgctg  
gaggattgtacgggtcctggaggattatacggacctggaggagctggagttcccggagcg  
ccaggagcttctggttagagcaggaggtatcggaggtgcagctggagctggtggtgctcg  
acctggtggagctctctggagggtgctggcggtatcaggtgtatcagttacagaatcagtta  
ctgttggtggagccggaggagcaggagctggtggaatcgggtggaccatcaggtctggga  
ggagccggagcaactggtggattcgggtggtcggggaggacctggtggacctggaggacc  
cgggtggaccaggaagatttgagggtgcagctggaggagcaggagccggtggtgctcgga  
ctggaggagctctctggaggcgctggcggtgctggcggtatcaggtgtaacagttgttagag  
tcagttagtggtggtggagccggaggaccaggagctggtggtgctgggtcctggagggtg  
cggacctggaggagttggaccgggaggtatttacggaccaggaggagctggaggacttt  
atggaccgggtgcagggggagccttcggatcaggaggaggagctggtgcaccaggaggg  
cctggagggtccaggtggtccaggcggcccaggaggtcctggaggaggagtaggcggagc  
aggaaccggcggtggtgttgcccaggaggttgagggtgttgtagacctctggaggggag  
gtggcaccgggtccggtatctgtctcttcaactataacagtcgggtggaggccaatcttca  
ggtggtgttttaccttcgaccagttatgtccaacgacaagcggtatgaaagattacc  
aaatttgattaatggtattaagagctccatgcaaggaggtggttaattatcagaatt  
ttggaaacattctgtcgcaatatgccacaggttctggaacatgcaactattatgatata  
aatcttttgatggatgcccttttggcgcgcttcacaccctcaactaccaggaggacctc

ttatgttccatcatacccttcgccctctgaaatgttatcgtacacggaaaatgttcgaa  
gatacttctga

SEQ ID NO: 18

Argiope trifasciata flagelliform silk protein (Flag) mRNA,  
partial cds. (without C-term.)

Genbank Accession: AF350265

DNA sequence (3008 bp)

cggcgcaccaggaggaggccctggcgggtgctggaccaggtggagcaggatttggtcctg  
gaggtggagctggatttggtcctggaggtggagctggatttggtcctggaggagcagca  
ggaggtcccgtggtccaggaggtccaggcggcccaggaggagccggaggttatggacc  
aggtggagccggcggttatggaccaggaggagtcggaccaggtggtgccggaggttatg  
gaccaggtggagccggaggttatggacctggaggatccggaccaggtggtgcaggacca  
ggcgggtgccggaggcgaggggtcccgtaacagtggatgtggacgtaactgttggacctga  
aggagtgggtggaggacctggcgggtgctggaccaggtggagcaggatttggtcctggag  
gtggagctggatttggtccgggaggagcacctggagcgccaggaggtcccgggtggtcca  
ggaggcccaggaggtccaggcggaccggaggagtcggacctggaggagccggaggtta  
tggaccaggtggagccggaggtgttggaccagctggaactggaggttttggaccaggtg  
gagccggaggttttggaccaggtggagccggaggttttggaccaggtggagctggaggt  
tttggaccaggtggagctggaggttatggaccaggaggagtcggaccaggtggagccgg  
agggtttggacctggaggagtcggaccgggtggttcaggacctggcgggtgcaggaggcg  
aggggtcccgtaacagtggatgtcgacgtaagtgttggcggcgccaccaggaggaggccct  
ggcgggtgctggaccaggtggagcaggatttggtcctggaggtggagctggatttggtcc  
tggaggtggagctggatttggtcctggaggagcagcaggaggtcccgtggtccaggag  
gtccaggcggcccaggaggagccggaggttatggaccaggtggagccggcggttatgga  
ccaggaggagtcggaccaggtggtgccggaggttatggaccaggtggagccggaggtta  
tggacctggaggatccggaccaggtggtgcaggaccaggcgggtgccggaggcgagggtc  
ccgtaacagtggatgtggacgtaactgttggacctgaaggagtgggtggaggacctggc  
ggtgctggaccaggtggagcaggatttggtcctggaggtggagctggatttggtccggg  
aggagcacctggagcgccaggaggtcccgtggtccaggaggcccaggaggtccaggcg  
gaccgggaggagtcggacctggaggagccggaggttatggaccaggtggagccggaggt  
gttggaccagctggaactggaggttttggaccaggtggagccggaggttttggaccagg  
tggagccggaggttttggaccaggtggagctggaggttttggaccagctggagctggag  
gttatggaccaggaggagtcggaccaggtggagccggagggtttggacctggaggagtc  
ggaccgggtggttcaggacctggcgggtgcaggaggcgaggggtcccgtaacagtggatgt  
cgacgtaagtgttggcggcgccaccaggaggaggccctggcgggtgctggaccaggtggag  
caggatttggtcctggaggtggagctggatttggtcctggaggtggagctggatttggt  
cctggaggagcagcaggaggtcccgtggtccaggaggtccaggcggcccaggaggagc  
cggaggttatggaccaggtggagccggcggttatggaccaggaggagtcggaccaggtg  
gtgccggaggttatggaccaggtggagccggaggttatggacctggaggatccggacca  
ggtggtgcaggaccaggcgggtgccggaggcgaggggtcccgtaacagtggatgtggacgt  
aactgttggacctgaaggagtgggtggaggacctggcgggtgctggaccaggtggagcag  
gatttggtcctggaggtggagctggatttggtccgggaggagcacctggagcaccagga  
ggtcccgtggtccaggaggcccaggaggtccaggcggaccggaggagtcggacctgg  
aggagccggaggttatggaccaggtggagccggaggttttggaccaggtggaactggag  
gttttggaccaggtggagccggaggttttggaccaggtggagccggaggttttggacca

ggtggagctggaggttttggaccaggtggagccggaggttatggaccaggaggagttgg  
accaggtggagccggagggtttggacctggaggagtcggacccggtggttcaggaccag  
gcggtgcaggaggcgaggggtcccgtaacagtggatgtcgacgtaagtgttggcggcgca  
ccaggaggaggccctggcggtgctggaccaggtggagcaggatttggctcctggaggtgg  
agctggatttggctcctggaggtggagctggatttggctcctggaggagcagcaggaggtc  
ccagtgggtccaggaggtccaggcggcccaggaggagccggaggttatggaccaggtgga  
gccggcggttatggaccaggaggagtcggaccaggtggtgccggaggttatggaccagg  
tggagccggaggttatggacctggaggatccggaccaggtggtgcaggaccaggcgggtg  
ccggaggcgaggggtcccgtaacagtggatgtggacgtaactggttgacctgaaggagtg  
ggtggaggacctggcggtgctggaccaggtggagcaggatttggctcctggaggtggagc  
tggatttggctccgggaggagcacctggagcgccaggaggtcccggtggtccaggaggcc  
caggaggtccaggcggaccggaggagtcggacctggaggagccggaggttatggacca  
ggtggagccggaggtggtggaccagctggaactggaggttttggaccaggtggagccg

SEQ ID NO: 19

*Dolomedes tenebrosus* fibroin 1 mRNA, partial cds.

Genbank Accession: AF350269

DNA sequence (2565 bp)

ctggttctggacaaggcagatacgggtggtcaaggtagttcaggaggctatggacaaggt  
gctggagctggagctgccaccgcccgaactgctagggctgatggatcgggacaaggccg  
atacgacggtcaaagtagtcaaggaggttatggacaaggtgctggtgctggagctaccg  
ccacggctgctgctgggggagctggttctggacaaggtggatatggtggccaaggtggt  
cttggaggctatggtcaaggagctggtgctggagctgcagccgctactgcagctggtgg  
agctggatccggacaaggtgattacggtgatcaaggtggtctaggaggatatggtcaag  
ggttctggagctggttctgcaaccgctcctgctgctggtggatctgggtttggacaaggg  
ggtttcggtaatcgaggtggaaaaggagcttatggtcaaagtgctggagctggagttgg  
agctgccgccaccgctgctgctggtggagctggttccggacaaggcggatacgggtgatc  
aaggtggtctaggaggatatggtcaaggtgctggagctggtgctgcctccgctgctgct  
ggaggtggagatggatacgaacaaggtggatacggtaatacaaggtggttttaggaagttt  
tgggtcaaggagctggggctggagctgccgccgacgcttctgctggtggagctggttccg  
gacgaggcggatacgggtgatcaaggtggtctaggaggatatggtcaaggtgctggagct  
ggtgctgcctccgctgctgctggaggtggagatggatacggacaaggttattacggtga  
tcaaggtggtcgaggaggatatggtcaaggttctggagctggttctgcaaccgctgctg  
ctgctggtggagctgggttttgacaaggcggatacggacaaggtggatacggtaataca  
ggtggttttaggaagttttggtcaaggagctggggctggagctgccgccgcccgttctgc  
tgggtggagctggttccggacgaggcggatacgggtgatcaaggtggtctaggaggatatg  
gtcaaggtgctggagctggtgctgccgccgctgctgctggaggtggagatggatacgg  
caaggtggatacggtaatacaaggtggttttaggaagttttggtcaaggagctggggctgg  
agctgccgccgcccgttctgctggtggagctggttccggacgaggcggatacggacaag  
gtggatacggtaatacaaggtggttttaggaagttttggtcaaggagctggggctggagct  
gccgccgcccgttctgctggtggagctggttccggacgaggcggatacgggtgatcaagg  
tgggtctaggaggatatggtcaaggtgctggatctggtgctgccgccgctgctgctggag  
gtggagatggatacggacaaggtggatacggtaatacaaggtggttttaggaagttttggt  
caaggagctggggctggagctgccgccgcccgttctgctggtggagctggttccggacg  
aggcggatacggacaaggtggatacggtaatacaaggtggttttaggaagttttggtcaag  
gagctggggctggagctgccgccgcccgttctgctggtggagctggttccggacgaggc

ggatacgggtgatcaaggtgggtctaggaggatatgggtcaaggtgctggagctgggtgctgc  
ttccgctgctgctggaggtggagatggatacggacaaggtggatacggtaatacaacgtg  
gtgtaggaagttatgggtcaaggagctggggctggagctgccgccacctctgctgctggt  
ggagctgggttccggacgaggcggatacgggtgaacaaggtgggtctaggaggatatgggtca  
aggtgctggagctgggtgctgcctccactgctgctggaggtggagatggatacggacaag  
gtggatacggtaatacaaggtgggtcgaggaagttatgggtcaaggatctggggctggagct  
ggagctgccgtagccgctgctgctgggtggagctgtttcgggacaagggggatacgtatgg  
tgaaggtgggtcaaggaggatatgggtcaaggttctggagctggagctgccgttgccgctg  
cttctgggtggaaccggagccggacaaggcggatacgggaagccaaggtagtcaagctggt  
tatgggtcaaggtgctggatttagagctgcagccgccaccgccgctgctgggtgctgggtg  
cgccggaggcggacaagggggatacggaggtcaaggaggttatgggtcaaggaaactgggtg  
ctgggtggtgctagttctgctggactttctgttactgtgggcaacatgggtttctcgtctt  
tcttctcccgaagctgcttctagagtttcttcggcagtttctagcttgggtgtcaaattgg  
tcaagtaaatgttgatgcattgcctagattattttcaaattcttcttctctatcagtg  
catctgctacaactgcttccgattgtgaggtcttgggtcaagttcttctagaggtgggtg  
tcagctctcgtgcaaatcgtctgctcg

SEQ ID NO: 20

Dolomedes tenebrosus fibroin 2 mRNA, partial cds.

Genbank Accession: AF350270

DNA sequence (2078 bp)

gttatgggtcaaggttctggagctgggtgctgccgctgctgctgccgctgctgggtgggtgct  
ggacaaagtgggttcaggtccttatgggtgcaagttatctatcaagcacaacatatataaac  
atcatcacaaggagcaggagggcggagtaggcggttacgggcaaggtagtggaaacgggat  
ctgcagctgcagctgctgggtgctgctggagcaggacaaggcggacaaggcgggttacgga  
caaggtgctggacaaggtgggtctaggaggttatgggtcaaggtgggtggagctgggtgctgc  
cgctgccgcagccgccgctgctgggtggagctggatctgggtcaaggtggatatgggtgggtc  
aaggtgggtctaggaggttatgggtcaaggtgctggagctggggctgcagccgccgctgct  
gctgggtggagccggagccggacaaggtgggttctcgggtgggtcaaggaggggtatggccaagg  
tggtggagctgggtgctgccgctgccgctgccgcagccgccgctgctgggtggagctggat  
ccggtcaaggtggatatgggtgggtcaaggtgggtctaggaggttatgggtcaaggtgctgga  
gctggggctgccgccgccgctgctgctgggtggagctggagccggacaaggtagttaacgg  
tggaacaaggaggttatggacaaggtggagctgggtgctgccactgccactgccgccgccg  
ctgggtggagctggatccggtcaaggtggatatgggtgggtcaaggtggactaggaggttat  
gggtcaaggtgctggagctggagctgctgccgccgctgccgctgctgctgggtggagccgg  
tgccggacaaggtggatacggaggtcaaggtgggtcaaggaggttatggccaaggtgctg  
gagctggagccgccgccgctgctgctgggtggagccggagccggacaaggtgggttacggt  
gggtcaaggaggttatggccaaggtgggtggagctgggtgctgccgctgccgcagccgccgc  
ttctgggggatctggatccggtcaaggtggatatgggtgggtcaaggtgggtctaggaggtt  
atgggtcaaggtgctggagctggggctggagctgctgcttctgctgctgctgctggagct  
ggatctggacaaggtggatatgggtggacaaggtgggtcttggaggttacggtcaaggagc  
tgagctggagctgctgctgggtgcttctgggttctgggttctgggtgctggacaaggtg  
gattaggaggatatgggtcaaggtgcaggagcaggtgccgctgctgcagctgctgggtgct  
agtggggcaggacaaggcggatttgggtccctatgggttctagttaccaatcaagcacctc  
atattcagtaacatcacaaggtgctgctgggtggattaggaggatatggacaaggtagtg  
gagctgggtgctgcagctgcaggtgccgcaggacaaggtgggtcaaggtgggttacgggtcaa

ggtgctggggcaggagccggggctggtgctggacaaggtggattaggaggatacgggtca  
aggtgctggttcttccgccgcttctgctgcggctgctggtggagctggagcaggacaag  
gtggatacgggtggtcaaggtggtctgggtggttacgggtcaaggtgctggagctggagct  
tccgccgccgcatctgctagtggagccggttctggacaaggtggatacggaggtcaagg  
aggttacggccaaggaactggtgctggtgctgctagtctgctggagttgctgttactg  
tgggcaacacgggttctcgtcttcttctccccaagctgcttctagagtttctctcagca  
gtttctagcttggtgtcaaatggtcaagtaaatgttgctgcattgcctagtattatttc  
aagcctctcttcttctatcagtgcattcttctacagctgcttccgattgtgaggtcttg  
tccaagttctgcttgagatcgtgtcggtcttctgtgcaaatcgtcagctcggccaacgtt  
ggatatattaatcctgaagcttccggttctctaaacgctgtcggatctgccttggcagc  
cgcaatgggttga

SEQ ID NO: 21

*Euagrus chisoseus* fibroin 1 mRNA, partial cds.

Genbank Accession: AF350271

DNA sequence (2207 bp)

gtaacgctagtcagattgcagcaagcgtagcatcagcggtcgcttcgagcgcattccgcg  
gcggcagccgctgcctcttctcctcagcagcagcagctgcaggcgccagttcggtgcgg  
tgctgcttcgagctcttcaacgactactactacaagtacctcctcgtctgcagcagccg  
cggccgcagcagcggcagcagcttcagcttcaggagcatcgagtgcctcggcagcagcc  
tccgcatcggcagcagctagcgccttctcttcagctctgatcagcagctttttgggaat  
aggagttttcggtaacacctttggttccatcgggtcggcgtcagctgccagttcaattg  
catcagccgctgctcaggcagcgccttctggacttggtttaagctatctcgttccagc  
ggagctagtgcagtagccagcgcagctgcaggggtcgggtgttgagctggagcatacgc  
ttacgcatacgtattgcaaattgcattcgcattccatactggcaaacacagggttactga  
gcgtgtcttcagcagcttcggttgcgagtagtggtggttccgctatcgctaccagcgtt  
tcctcttcgtccgccgcagcagcagcatcagccagtgagcagcagcagcatcagccag  
tgagcatcgtcagcatcggcaagcagcagtgcatcagcagctgccgcagccggggtt  
ccgcggccgctggagctgcttcgtcggcatctgcttccgcagcagcgtctgccttcagc  
tcggctttcatctcagctttacttggtttctcacaatttaacagcgtcttcggttccat  
tacctccgcgtcactcggacttggcatcgcagcgaacgctgttcagtcgggacttgcat  
cccttggtctaggagctgcggcttcggcagcagcatctgcagtggaacgcagggtta  
aacggctctggtgcataatgcttacgcgacagctattgcctcggcgataggaaacgcact  
tcttggtgccggattcttgacagctggtaacgctagtcagattgcagcaagcgtagcat  
cagcggtcgcttcgagcgcattccgcggcggcagccgctgcctcttctcagcagcagct  
gcaggcgccagttcggtgcgggtgctgcttcgagctcttcaacgactactactacaag  
tacctcctcgtctgcagcagccgcggccgcagcagcggcagcagcttcagcttcaggag  
catcgagtgcctcggcagcagcctccgcattcggcagcagctagcgccttctcttcagct  
ctgatcggcgatcttttgggaataggagttttcggtaacacctttggttccatcgggtc  
ggcgtcagctgccagttcaattgcattcagccgctgctcaggcagcgtttctggacttg  
gtttaagctatctcgttccagcgggagctagtgcagtagccagtgagtcgaggggtc  
ggtgttgagctggagcatacgttacgcatacgtattgcaaattgcattcgcattccat  
actggcaaacacagggttactgagcgtgtcttccgcagcttcggttgcgagtagtggtg  
cttccgctatcgctaccagcgttctcttctcgtccgccgcagcagcagcatcagccag  
gcagcagcagcagcatcagccggtgcatcagcagcatcgtcagcatcggcaagcagcag

tgcatacagcagctgctggagcaggtgctggagcaggtgctggagcttcaggtgccagtg  
gagctgcaggaggatcaggtggcttcggtttatcgtctggtttcgggtgctggaatagga  
ggtttagtggtgggtacccctctggcgcgctgggaggtcttggtattccttctggtttgc  
ctcatctggtttattgtctccagctgcaaatacaagaattgcttctctgatccctttga  
ttttgtctgcgatttcaccgaatggcgtaaactttggtgtgattggaagtaatttgca  
tcttttagcttcgcaaatactcaaagtgggtggaggtattgcagcgtctcaagcttttac  
ccaagctttgctggaattagtcgctgcctttattcaagttctgtcttctgctcaaatacg  
gtgcagtttagtagctcttcagcaagcgcaggcgctactgccaacgcatttgctcaatcg  
ctgtcgtcagctttttgcgggatag

SEQ ID NO: 22

Plectreureys tristis fibroin 1 mRNA, partial cds.

Genbank Accession: AF350281

DNA sequence (2740 bp)

cgccgccgcccgcagctgcagcagcagccgggtgccggagcaggggctggagcaggag  
caggtgctggagcaggagcaggatctggagcttcacatcgggtctctaccagttcatcg  
agcggatccggagcaggtgcaggagcaggttctggagctggatctggcgaggagcagg  
ttctggggcaggtgcaggagcaggcgctggtggtgcaggagcaggtttcggcagtgcc  
tcggattaggctatggagtaggattgtctagtgcacaagcgcaggcacaggcccaagct  
gccgcgaggcacaaagcacaggctcaggcccaggcatacgcagcagcacaaagcacaggc  
acaagcacaaagcacaaagcacagcccgccgcccgcagctgcagcagcagccgggtg  
ccggagcaggggctggtgcaggagcaggtgctggagcaggagcaggatctggagcttc  
acatcgggtctctaccagttcatcgagcggatccggagcaggtgcaggagcaggttctgg  
agctggatctggcgaggagcaggttctggggcaggcgaggagcaggcgctggtggtg  
caggagcaggtttcggcagtggtctcggattaggctatggagtaggattgtctagtgc  
caagcgcaggcacaggcccaagctgccgcgaggcacaaagcacaggctcaggcccaggc  
atacgcagcagcacaaagcacaggcacaaagcacaaagcacaaagcccgccgcccgcg  
ccgcagctgcagcagcagccgggtgccggagcaggggctggagcaggagcaggtgctgga  
gcaggagcaggatctggagcttcacatcgggtctctaccagttcatcgagcggatccgg  
agcaggtgcaggagcaggttctggagctggatctggcgaggagcagggctctggggcag  
gcgaggagcaggcgctggtggtgcaggagcagctttcggcagtgccctcggattaggc  
tatggagtaggattgtctagtgcacaagcgcaggcacaggcccaagctgccgcgaggc  
acaagccgacgctcaggcccaggcatacgcagcagcacaaagcacaggcacaaagcacaa  
cacaaagcacaaagcccgccgcccgcgcccgcagctgcagcagcagccggtgccggagcagg  
gctggtgcaggatcaggtgctggagcaggagcaggatctggagcttcacatcgggtctc  
taccagttcatcgagcggatccggagcaggtgcaggagcaggttctggagctggatctg  
gcgaggagcaggttctggggcaggcgaggagcaggcgctggtggtgcaggagcaggt  
ttcggcagtgccctcggattaggctatggagtaggattgtctagtgcacaagcgcaggc  
acaggcccaagctgccgcgaggcacaaagccgacgctcaggcccaggcatacgcagcag  
cacaaagcacaggcacaaagcacaaagcacaaagcccgccgcccgcgcccgcagctgca  
gcagcagccggtgccggagcaggggctggtgcaggatcaggtgctggagcaggagcagg  
atctggagcttcacatcgggtctctaccagttcatcgagcggatccggagcaggtgcag  
gagcaggttctggagctggatctggcgaggagcaggttctggggcaggcgaggagca  
ggagctggtggtgcaggagcaggtttcggcagtgccctcggattaggctatggagtagg  
attgtctagtgcacaagcgcaggcacagtcagcagctgccgcagggcacaaagctgacg  
ctcaggcccaggcatacgcagcagcacaaagcacaggcacaaagcacaaagcacaa

gccgccgccgccgccgcagctgcagcagcagccggtgccggagcaggggctggtgcagg  
agcaggagctggagcaggagcaggatctggagcttccacatcgggtctctaccagttcat  
cgagcgcacccggagcaggtgcaggagcaggttctggagctggatctggcgcaggagca  
ggttctggggcaggcgcaggagcaggcgtggtggtgcaggagcaggtttcggcagtgg  
cctcggattaggctatggagtaggattgtctagtgcacaagcgcaggcacaggcccaag  
ctgccgcgcaggcacaaagcacaggctcaggcccaggcattagcagcagcacaaagcaca  
gcacaggcacaaagcacaaagcacaaagccgccgcagcgcagccgctgcagcagcagccgg  
tgccggagcaggggctggttcaggcgcaggagctggagcaggagcaggagcagggctctg  
gagcttccacatcgggtctctaccagttcatcgcagcgcagccggagcaggtgcaggagca  
ggttctggagccggagctggatctgggacaggcgcaggtattgctcttcccttcgatcgt  
tctctccctgcagcatcatcagaatttcgtccgtttcgtcctccgtccagtcagcag  
gttcgggtctcagtttctcctcgtgtcaaacacattgtcgcagacagcatcggctata  
agaagcagcaatcctcaactctcttcagcagatgttctgatccagagcttggtcgaaat  
catcgtcgggttggtacaagcgttactggttcttcagcgtcagcccaaactttcgtga  
actcattgtctcaggttgccgggttaa

SEQ ID NO: 23

*Plectreurys tristis* fibroin 2 mRNA, partial cds.

Genbank Accession: AF350282

DNA sequence (2293 bp)

gtacagattctgtcgcacccctcagcctctagctcggcgagtgcacccctcatcagcaacag  
ggcctgacacgggttatccagtagggactacggagcaggacaagcagaagcagcagcat  
cagcagcggcgggcggcgaggcagcaagcgcagcagaagcagcaacaattgcagggttgggct  
acggaagacaaggtcaaggtactgattctagtgcacccctcagtcctctacttcgacaagt  
taccctcattagcaacagggcctggctcgagatatccagtaagggaactacggagcagatc  
aagcagaagcagcagcatcagcagcggcgggcagcaagcgcagcagaagaatcgcaagct  
tggttacaggacgacaaggtcaaggtacagattctgtcgcacccctcagcctctagctcgg  
cgagtgcacccctcatcagcaacagggcctgacacgggttatccagtagggtaactacggag  
caggacaagcagaagcagcagcatcagcagcggcgggcgggcagcaagcgcagcagaag  
cagcaacaattgcagggttgggctacggaagacaaggtcaaggtactgattctagtgcac  
cctcagtcctctacttcgacaagtgtaccctcatcagcaacagggcctgacacgggttatc  
cagtagggtaactacggagcaggacaagcagaagcagcagcatcagcagcggcgggcggg  
cagcaagcgcagcagaagcagcaacaattgcagggttgggctacggaagacaaggtcaag  
gtactgattctagtgcacccctcagtcctctacttcgacaagtgtaccctcatcagcaacag  
ggcctgacatgggttatccagtagggaactacggagcaggacaagcagaagcagcagcat  
cagcagcggcgggcgggcagcaagcgcagcagaagcagcaacaattgcaagttgggct  
acggaagacaaggtcaaggtactgattctagtgcacccctcagtcctctacttcgacaagt  
taccctcatcagcaacagggcctggctcgagatatccagtaagggaactacggagcagatc  
aagcagaagcagcagcatcagcagcggcgggcgggcgggcgggcagcaagcgcagcagaag  
aaatcgcaagcttgggctacggacgacaaggtcaaggtacagattctgtcgcacccctcag  
cctctagctcggcgagtgcacccctcatcagcaacagggcctgacacgggttatccagtag  
ggtactacggagcaggacaagcagaagcagcagcatcagcagcggcgggcgggcagcaa  
gcgcagcagaagcagcaacaattgcagggttgggctacggaagacaaggtcaaggtactg  
attctagtgcacccctcagtcctctacttcgacaagtgtaccctcatcagcaacagggcctg  
gtcgcagatatccagtaagggaactacggagcagatcaagcagaagcagcagcatcagcaa  
cggcggcgggcgggcgggcagcaagcgcagcagaagaatcgcaagcttgggctacggac

gacaaggtcaaggtacagattctgtcgcatcctcagcctctagctcggcgagtgcatcct  
catcagcaacagggcctgacacgggttatccagtagggtagctacggagcaggacaagcag  
aagcagcagcatcagcagcggcgggcggcagcaagcgcagcagaagcagcaacaattg  
caggtttgggtacggaagacaaggtcaaggtactgattctagtgcacccagtcctcta  
cttcgacaagtgatcctcatcagcaacagggcctggctcgagatatccagtaattggact  
acggagcagatcaagcagaagcagcagcatcggcagcggcggcggcgagcagaagcag  
caacaattgcaggtttggactacgaaggacaaggacaaggtactgattctgggtgcaccc  
cagtttctagttcgacaagtgatcctcatcagcaacaggtgttactcaaactacgatcg  
cccttccccctgacgtatccgcacgaatctcgttcctcacgtcatatcttgcatccgcag  
gttcaggtctcagcctctacacgctatccaacctactgtcgagacagcgttggccataa  
gcaagagccgtcctgaactctctcccaacgaagtcctaattcaaagtttagctgagatca  
tagtggttttggtacaagcgtcactaaacaagccagctcttcggcatcgggtgcaatatt  
tcgggcgtttcct

SEQ ID NO: 24

*Plectreurys tristis* fibroin 3 mRNA, partial cds.

Genbank Accession: AF350283

DNA sequence (6052bp)

cgcaatcagctcgagtttgtacgctttcaattaccaggcgtcggcggaagttcagctgc  
tgacagagctcgcccaaactgcgtctacttcagcaaacagacagctgcaagtacgtc  
tgcatcaacagcagcaacttctacaacacagacagctgcaacaacgtctgcacggc  
agcaagttcacaaacggttcagaaagcaagcagagttccgccgcacaaactgctgcctc  
caagtctcagagcagttccgcgggcagttcgagaacgacctcaactgctgcagcatccgc  
aagcagcagttatgcattcgacaaaagttatcgagctatctcctgtcttcgcagcaatt  
cacgactgccttcgcaagttctaccgccgtagcgtcctctcagcagtagcgggaagccat  
ggcccagctctgtcgccacgtctcttggaactcggctacacatatgcgtctgcactttctgt  
cgccatggcacaagccatctccggggttgggcggaggagctagcgttacagttacgcaac  
ggccatttcgcaagccatttctagagctcttacaagttccggcgatctctgtcctcttc  
gcaagcgacgtctgttgcttccgcaatcagctcgagtttgtacgctttcaattaccaggc  
gtcggcggaagttcagctgctgcacagagctcggcccaaactgcgtctacttcagcaaa  
acagacagctgcaagtacgtctgcacacagcagcaacttctacaacacagacagctgc  
aactacttctgcacgcggcggaagttcacaaacggttcagaaagcaagcacgagctc  
cgccgcacaaactgctgcctccaagttctcagagcagttccgtgggcagttcgacgacctc  
aactgctgcagcatccgcaagcagcagttatgcattcgacaaaagtttatcgagtatct  
cctgtcttcgcagcaattcacgactgccttcgcaagttctaccgccgtagcgtcctctca  
gcagtagcgggaagccatggcccagctctgtcgccacgtctcttggaactcgggtacacata  
tacgtctgcactttcggtcgccatggcacaagccatctccggggttgggcggaggagctag  
cgcttacagttacgcaacggccatttcgcaagccatttctagagttcttacaagttccgg  
catatctctgtcctcttcgcaagcgacgtctgttgcttccgcaatcagctcgagtttgta  
cgctttcaattaccaggcgtcggcggaagttcagctgctgcacagagctcggcccaaac  
tgctgtctacttcagcaaacagacagctgcaagtacgtctgcacacagcagcaacttc  
tacaacacagacagctgcaactacttctgcacgcggcggaagttcacaaacggttca  
gaaagcaagcacgagctccgccgcacaaactgctgcctccaagttctcagagcagttccgt  
gggcagttcgacgacctcaactgctgcagcatccgcaagcagcagttatgcattcgacaca  
aagtttatcgagtatctcctgtcttcgcagcaattcacgactgccttcgcaagttctac  
cgccgtagcgtcctctcagcagtagcgggaagccatggcccagctctgtcgccacgtctct



tggactcggctacacatatagctctgcactttctgtcgccatggcacaagccatctccgg  
ggttggcggaggagctagcgcttacagttacgcaacggccatttcgcaagccatttctag  
agttcttacaagttccggcgatatctctgtcctcttcgcaagcgacgtctgttgcttccgc  
aatcagctcgagtttgtagcgtttcaattaccgggctcggcggcaagttcagctgctgc  
acagagctcggcccaaactgcgtctacttcagcaaaacagacagctgcaagtacgtctgc  
atcaacagcagcaacgtctacaacacagacagctgcaactacttctgcatcgacggcagc  
aagttcacaacgggttcagaaagcaagcacgagctccgcccgtcaactgctgctcagca  
gactgggcaatcctcttctgtacagaatcaaggaagctcctctgccagctcgagttcagt  
cagcgtttctgatatctccgattctctcacaacatctttgctgcagctctgaagaattcac  
atcggccttcggaagcacagtttagcgaggctgaggccagtcgtacgaggaggccgtggc  
tcagtctactgtcgcaactcgggatatagattattctcaaagctccgctctcgtactgc  
tgtagcaaacgcagtatcacaagttaacaaggctccagttctcgcgttatgcccgcgc  
catagcatatgcaatcacgacgtacctgaaaactactcgaattattactactattactag  
aactcaagtgaatcatcttgctctgcaatcagttcgagcctgtctacagcgaggcgac  
atctagtgc aaatgcatatcaggaacagaccactcagttctctgcagcagcaagtgcggc  
agcacagttccagtgagttc aaacgcagaacactcagttctctgctctgcggcaagcag  
tgatgcaagcacttccctaccagacacagcagagttactcggacgcgtcggcagccagcgt  
tgcagcagaaagcacaaagcgcgaatcaagcgcagagcacgcagtcacggccgcgcaag  
cagctctacaaattctgcctaccagagccaacaaagctacatagatgcttctacggtcag  
ctctgcgtctgcaaatacagcgcagtcgacttaccaagtaacaattcctgataatacgt  
ttttgctgaatctctgtcatccacactgatacaacatgagcaattcaattcgaaattcgg  
aagctacattccactagtaactgctcgggagttatgcttcggcaatggctcgagcaacagc  
tcttatcattgggttttgacagcactggaacttcagcacttgagttctgcggtcgcagtagc  
ggatccaatgtcgattatgccagcgcatattcctacgccagagcaatagcatttgcaat  
tagcaatgtacttaccaacaatggaatattcgcgtcagcctcagaagcactatatcttgc  
ccctgccatgtagcaagtttgcatgcatttggtaagtcgagcttttctgaaagttcggc  
attcgcattgggctaacagcatctctccgtcaacagcaataacgtccgcgcaaagcagcag  
tgatctgctggcgcacatcttcaggacaaagctcatatgacactagcagtgctgttctctc  
agccagcagcgcagaagcaacgggaatcttcaagcgtctttgatacttatcaagctacgca  
aatcgaaagttctgcagccgcgcagccgcacgtcatcggcataatgactcgcaatttctc  
tgaatcttctctgctagcagtgacgcagcgttcagcttttctcggaacagacctcctatga  
cataagcagtgacttatcttcagcaagcagcgtactgccgcagctgcttcgtcctcagc  
ttatgaatcgcaattttcggacgcttcttcggtagcagtgagctgccgctgcttcttc  
gcagcagaactcatagcacaccgatgccttgattcagcaagcagcgtgcttcgctgc  
cgcatcggcctcagcttacgaattggaattttcggacgcttcttctagcagcagcgagc  
tgccgttgcttcttcgcagcagggctcatagcacacaagcagtgacttctcttcagcgag  
cagcgtgctgggcccagctgcatcggttacgaatcgaaatttttggaacgcttcttctag  
cagcagcgcagctgccgctgcttcttcgcagcagagctcatagcaaacagcagtgactt  
agtttcagcgagcagcgtgctgctgagctgcatcggcctcggttaccaatcgcaatt  
tttggaacgcttcttctagcagcaatgcagctgccactacttcttcgcggcagagctcata  
tgacacaagcagtgacttctcttcagccagcatcgctgcggccgcagctgcatcggcctc  
gtcttatgaatcgcaattttcggacgcttcttctagcagcaatgcggctgccgctgcttc  
ttcgcagcagagctcatagatacaagcagtgacttagtttcagctgcatcggcctcggc  
ttatgaatcgcaatttttggaacgcttcttctagcagcaatgcagctgccactacttcttc  
gcagcagagctcatatgacacaagcagtgacttctcttcagccagcatcgctgcggccgc  
agctgcatcagcctcgtcttatgaatcgcaattttcggacgcttcttctagcagcaatgc  
ggctgccgctgcttcttcgcagcagagctcatagatacaagcagtgacttagtttcagc

gagcagcgctgcggccgcagctgcatcgccctcgtcttatgaatcgcaattttcggacgc  
ttcttctagcagcaatgcggctgccgctgcttcttcgcagcagagctcatagatacaag  
cagtgaacttagtttcagcgagcagcgctgcggctgcagctgcatcgccctcggttatga  
atcgcaattttcggacgcttcttctagcagaaatgcggctgccgctgcttcttcgcagca  
gagctcatagatacaagcagtgacttagtttcagcgagcagcgctgcggccgcagctgc  
atcgccctcgtcttatgaatcgcaattttgagcgttcttctagcagcaatgcagctgc  
cactacttcttcgcagcagagctcatatgacacaagcagtgacttcttctcagccagcat  
cgctgcggccgtagctgcatcagcctcgtcttatgaatcgcaattttcggacgcttcttc  
tagcagcaaagcagctgccgctgcttcttcgcagcagagctcatagatacaagcagtg  
cttagtttcagcgagcagcgctgcggccgcagctgcatcgccctcgtcttatgaatcgca  
attttcggacgcttcttctagcagcaatgcagccgccgctgcttcttcgcagcagagctc  
atacgatacaagcagtgacttcttctcagcgaacagtgctgctctagcagaatcttcagc  
tgccactgaaatttaccagagacacaaatcgcaagttccattgcagccgcttcagcatt  
gtcgggaagcacatacgtcagaattggccgaagcttcttccagcagcagtgagcttctgc  
agcagcagcagcagcttcggaacaaagcctttacgacacgagcagtgccgcttcttcagc  
aagcagcagcagcttcatagcttcttcggatattccgtaataacagagtttgccgttaa  
ctctgcagcgagcagcagtgagcggaagagagcggttcgcaagttgacgaagaaacgta  
ccaaaactttgatcagtagcttctcaatttcagcgctcagcatcggcagctcagagctcaga  
aatttaccagatgtatcctcctcttcggcagcagcctctacatcttcagcagcgctcttc  
cttggaacatctggaacagttgcagaaagcggtctacagcagcaagcagcagctatgc  
agcagcagcagcagcgctcctcatcagcggttcgacgagctcgccctcattcctgtcagc  
ggacagcctgtcgtcctctttggcttctctgagaatttggtccttttctcctaagctgat  
gtcttcttgtactcaggtgatggtctcgacatcgcgagttctccgatgcagtatcttc  
catggtttctagtagtcaaaagttcaaatccaggtgtaagtgttctcagatactcacaga  
actgctcttcgaggtaatcgtagcttttggtcaagctctcacaaaatcgaaagttttcaac  
tatggagacggctgaatctctaatagcggccttcgcacaagctttcgtctaa

SEQ ID NO: 25

Plectreureys tristis fibroin 4 mRNA, partial cds.

Genbank Accession: AF350284

DNA sequence (5446bp)

gtcccagcaaggacctatcggaggtgtcggcggtcgaaacgccttttagtagttccttcgc  
cagcgcaactcagtttgaaccgaggatttaccgaagttatcagcagcgccctccgcaaccgc  
ggttgcttctgccttccagaaagggttgaccttacggtagcgcatcgccttgtctgc  
agcgagcgcagcagctgacgcttacaactctattgggttcaggagcaaatgcgttcgccta  
tgctcaagctttcgcaagagtactgtatccgttggttcgacaatatggtttatcttcgag  
cggtaaagcttctgcttctcgccagtgccattgcaagctcttttagcagtggtacctccgg  
ccaaggaccgtcgattggacaacaacagcctccggttacgatctcagcggcgtcagcatc  
agcaggtgcttcggctgctgccgtaggaggcgggcaagtaggcccaggggcatatggtgg  
gcagcaacagagcactgcagcttcggcttcagcagcagcagctactgctacttctggcgc  
agcacagaagcagccttcaggcgagtcctcagtggaactgcacggcggcagcaacttc  
ggttacttctggcgagcgccggttggaaccagaggttcagcaccataatatttatcc  
gcaaggtccattgcaacaaggaccagccccggaccttccaacgtccagccaggaacgtc  
ccagcaaggacctatcggaggtgtcggcggttcgaacgccttttagtagttccttcgccag  
cgcaactcagtttgaaccgaggatttaccgaagttatcagcagcgccctccgcaaccgcggt  
tgcttctgccttccagaaagggttgaccttacggtagcgcatcgccttctgtctgcagc

gagcgcagcagctgacgcttacaactctattgggttcaggagcaaatacggttcgcctatgc  
tcaagctttcgcaagagtactgtatccgttgggttcgacaatatggtttatcttcgagcgg  
taaagcttctgctttcgccagtgccattgcaagctcttttagcagtggtagctccggcca  
aggaccgctcgattggacaacaacagcctccggttacgatctcagcggcgctcagcatcagc  
aggtgcttcggctgctgccgtaggaggcgggcaagtagggccaggggcatatgggtgggca  
gcaacagagcactgcagcttcggcttcagcagcagcagctactgctacttctggcgcagc  
acagaagcagccttcaggcgagtcctcagtggaactgcacgcggcgagcaacttcgggt  
tacttctggcggagcgcgggttggaaccaggagttccagcaccaatatatttatccgca  
aggtccattgcaacaaggaccagccccggaccttccaacgtccagccaggaacgtccca  
gcaaggacctatcggaggtgtcggcgggtcgaacgccttttagtagttccttcgccagcgc  
actcagtttgaaaccgaggatttaccgaagttatcagcagcgcctccgcaaccgcgggtgc  
ttctgccttcagaaaagggttgaccttacgggtacggcattcgctttgtctgcagcgag  
cgcagcagctgacgcttacaactctattgggttcaggagcaaatacggttcgcctatgctca  
agctttcgcaagagtactgtatccgttgggttcaacaatatggtttatcttcgagcgctaa  
agcttctgctttcgccagtgccattgcaagctcttttagcagtggtagctccggccaagg  
accgtcgattggacaacaacagcctccggttacgatctcagcggcgctcagcatcagcagg  
tgcttcggctgctgccgtgggaggcgggcaagtaggtcagggggccatatgggtgggagca  
acagagcactgcagcttcggcttcagcagcagcagctactgctacttctggcggagcaca  
gaagcagccttcaggcgagtcctcagtggaactgcacgcggcgagcaacttcgggttac  
ttctgccggagcgcgggttggaaccaggagttccagcaccaatatatttatccgcaagg  
tccattgcaacaaggaccagccccggaccttccaacgtccagccaggaacgtcccagca  
aggacctatcggaggtgtcggcgggtcgaacgccttttagtagttccttcgccagcgcact  
cagtttgaaaccgaggatttaccgaagttatcagcagcgcctccgcaaccgcgggtgcttc  
tgccttccagaaaagggttgaccttacgggtacggcattcgctttgtctgcagcgagcgc  
agcagctgacgcttacaactctattgggttcaggagcaaatacggttcgcctatgctcaagc  
tttcgcaagagtactgtatccgttgggttcaacaatatggtttatcttcgagcgctaaagc  
ttctgctttcgccagtgccattgcaagctcttttagcagtggtagctccggccaaggacc  
gtcgaatggacaacaacagcctccggttacgatctcagcggcgctcagcatcagcaggtgc  
ttcggtgctgccgtgggaggcgggcaagtaagtcagggggccatatgggtgggagcaaca  
gagcactgcagcttcggcttcagcagcagcagctactgctacttctggcggagcacagaa  
gcagccttcaggcgagtcctcagtggaactgcacgcggcgagcaacttcgggttacttc  
tgccggagcgcgggttggaaccaggagttccagcaccaatatatttatccgcaagggtcc  
attgcaacaaggaccagccccggaccttccaacgtccagccaggaacgtcccagcaagg  
acctatcggaggtgtcggcgggtcgaacgccttttagtagttccttcgccagcgcactcag  
tttgaaaccgaggatttaccgaagttatcagcagcgcctccgcaaccgcgggtgcttctgc  
cttccagaaaagggttgaccttacgggtacggcattcgctttgtctgcagcgagcgcagc  
agctgacgcttacaactctattgggttcaggagcaaatacggttcgcctatgctcaagcttt  
cgcaagagtactgtatccgttgggttcaacaatatggtttatcttcgagcgctaaagcttc  
tgctttcgccagtgccattgcaagctcttttagcagtggtagctccggccaaggaccgtc  
gattggacaacaacagcctccggttacgatctcagcggcgctcagcatcagcaggtgcttc  
ggctgctgccgtgggaggcgggcaagtaggtcagggggccatatgggtgggagcaacagag  
cactgcagcttcggcttcagcagcagcagctactgctacttctggcggagcacagaagca  
gccttcaggcgagtcctcagtggaactgcacgcggcgagcaacttcgggttacttctgc  
cggagcgcgggttggaaccaggagttccagcaccaatatatttatccgcaagggtccatt  
gcaacaaggaccagctcccggaccttctacgtccagccagcaacgtcgcagcaaggacc  
tatcggaggtgccggcgggtcgaacgcatttagtagttccttcgccagcgcactcagtg  
gaaccgaggatttagcgaagttatcagcagcgcctccgcaaccgcgggtgcttctgcctt

ccagaaagggcttgccccctacgggtacggcatttgcgtttatctgcagcgagcgctgcagc  
tgacgcttacaactctattgggttcaggagcaaatgcgttcgcctatgctcaagctttcgc  
aagagtactgtatccgttggttcaacaatatggtttatcttcgagcgctaaagcttctgc  
tttcgccagtgccattgcaagctctttcagcagtgggcgccgcccgaaggacagtcgat  
accatacgggtggacaacaacaacctccaatgacgatctcagcggcgctcagcatcagcagg  
tgcttcagctgctgccgtgaaaggcgggcaagtaggtcaggggcatatggtggccagca  
acagagcactgcagcttcggcttcagcagccgcaactactgctactgctggcggagccca  
gaagcacccttcaggcgaatactcagtggaactgcatcggcggcagcaacttcggttac  
ttctggcggagcgccggttggaaaaccaggagttccagcgccaatatcttatccgcaagg  
tccattgcaacaaggaccagcccccgaccttccaacgtccagccaggaacgtcgcagca  
aggacctatcggagggtgctggcgagtcgaacaccttagtagttccttcgccagcgact  
cgggtgggaaccgaggatttagcggagttatcagcagcgccctccgcaaccgcggtcgcgtc  
tgcttccagaaagggcttgccccctacgggtaccgcattcgtttatctgcagccagcg  
tgagctgacgcttacaactctattgggttcaggagcaagtgcgtctgcctatgctcaagc  
tttcgcaagagtgcgtgtaacctatgctccagcaatatggtttatcttcgagcgctgacgc  
ttccgctttcgccagtgctattgcaagttcttttagcactggggctcgccggccaaggacc  
gtcgggtaccatacgttggacaacaacagccttcgattatggtctcagcagcttcagcatc  
agcagctgcttcagccgctgccgtgggaggcgcccgagtagttcaggggcatatcgatgg  
aggacagcctcagcaaccgaacattgctgcttcggcggcagcagcagctactgctacttc  
tagtggaacctaaagaggagcctttgggcgagtcctctgtgatagctacatcgggtgacgc  
cgctcgtcggtttcttctggcggagccccagggtgtacaaggcggaggtccagtgacagt  
gtcttatcgtgaaggtccttctcaaatctcttcaacaacgctgctgcaggcggtacc  
ttctacgcagctctgttgggtctggtgttctggttgggcctaatcagtatgaaatggttta  
tgctcctttgcagcaattcgggtggtgttctcggttctaatttgccttcaccgtcggcaca  
tagcagaatagcatcgctgatgtcggacgtacttagtcttttttcgccaggaaactctgg  
ttttaactatggggggttttgctagagctctctcgtctgtggtcgcgcagttagccagtc  
taatgccaaagtgtcgaccactgacgttatcatcaagttttgatggaagctctagttgc  
gctaattgagctcttgcgggtgcgaaaattggtgtgttcatcccggtcggggtcaggc  
tggtgccagtgcttttgctcaacatttcggcagtgcggttggtga

SEQ ID NO: 26

Phidippus audax fibroin 1 mRNA, partial cds.

DNA sequence (1711 bp)

ggagctggagctggcgctggctatggtgcaggtgctggttcaggagctggtgcaggctc  
tggtgcaggagctggagcaggagctggagcaggagctggagcaggctatggagcaggag  
caggttcaggagctggtgctggcgaggttacggacgaggtgcaggagcaggagcggga  
gctggagcaggttacggccaaggtgctggagcgggagctggtgccggcgaggttatgg  
cgctggagctggatctggagctggagccggctatggtacaggtgctggttcaggagctg  
gttcaggagctggttcaggagctggatcaggagctggagcaggagctggagcaggtgca  
ggttatggagcaggagcaggttcaggagctggtgctggcgaggttacggacgaggtgc  
aggagcaggagcgggagctggagcaggttacggccaaggtgctggagcaggagctggtg  
ccggtgcaggtgctggttcaggagctggtgcaggttctggtgcaggagctggtgctggt  
gcaggttacggacaaggtgcaggagcaggagctggtgccggtgcagggtatggcgctgg  
agcaggttctggagctggagctggcgctggctacggtgcaggtgctggttcaggagctg  
gtgcaggttctggtgcaggagctggtgctggagcaggttacggtcaaggtgctggagct  
ggagctggcgccggtatggtgcaggtgctggttctggagctggtgcaggctctggtgc

aggagctggatcaggagctggagctgggttcaggctatggcgaggagctgggttcaggag  
ctggcgctggcgagggttatggacaagggtgccggagcagggtgctgggtgcagggtgcaggc  
tatgggtgcaggagcagggttctggagctggaactgggtgcaggctatgggtgctgggtgcagg  
tgcaggatatgggtgctgggtgcagggtgcaggagctgggttcaggagcagggtgccggggcag  
gttatgggtgctgggtgctgggtgcaggcgctggagcaggctatgggtgctggagctgggtcc  
ggascagggtgcaggarcagggttatgggtgctgggtgcagggtgcagggttcagggtgtaggagc  
agggtgctggagctgggtgctggagcaggatatggagctggagcagggtgcaggagcagggt  
atgggtgctgggtgcagggtgcagggtgctgggtgctgggtgcaggagcaggatatggcgctgga  
gcagggtgcagggtgcttctgtaagttccactgtatctaactgcttccagaatgtcttc  
agagaatacatcacgtcgtgttcttccagccatttcaagcattgtcggctctggtggag  
ttaacatgaattctcttcaaacgtaattcttaattgtatcatcgagcgttgcctgcattc  
aatcctggactgtctggatgtgaagttcttgggtcaaaccctgttgggaagtagtatctgc  
attgggtcacattttgagctatgcaagtgtgggtagtgttgatgccagcgtgctgggtc  
agtcagcccagactgtagctacagccatgagtagtgaatgggttgaattactttgacc  
tttcaatatttttgaagactttatgttgttactttttgaattacgtaattgtctgaaaaa  
taagataaataaatagaagtatatatgcnaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

SEQ ID NO: 27

Zorocrates sp. fibroin 1 mRNA, partial cds.

DNA sequence (845 bp)

gggtgcagcagccgcagcctcagcagcagcagcaggcggacgaggaagccaaggaggtta  
cggagatgacgggtgggtgcagcagcagcagcagcagcagcggcgagcagccgcggcag  
gaagtgggtggaaccggaggaggacaagggggcgcgagatggaggtgcggcagcagca  
gcagcagcagccgcagaggccgcagcagggtggaaaaggaagacaaggaagttacggaga  
tgacgggtgggtgcagcagtagcagcagcagctgcagcggcagcagcggcaggaagaggtg  
gttccggaagaggacaaggacttcgtagagataaaggaagttacggagttgacgggtgg  
gcagaagcagcagcatccgcagcgccacagcaggcagacaaggaagacaaggaagttta  
cggagatgacgggtgggtgcagcagcagcagcagcagcagcggccttctgcttcacgggttag  
cctcctcttctgctgtttctcgagtctcatctgctgtttctgctgttgcattgcaatggc  
ttttctgatgtaaattccctctccaacgtgatttctggactttctgcttctgtatcttc  
ttccacacctgagctgactggttgcaagttctcgtggaagtccttttgaagtagtat  
cagctttggttcataattttgaactttgctgacatttgaaacgttaataattagtgcttca  
gggtgattccacatccctttagggcgaactgttttagaagcctttggctgaaatattac  
tctattcctttttttttttgaatatgtttcagcttttaactgtgacataaaaaatgt  
tatataagggaataaatata

SEQ ID NO: 28

Argiope trifasciata aciniform fibroin 1 mRNA, partial cds.

DNA sequence of 3' region (709 bp)

ggcatcaatgtagatagcggcagtgtaaaagtgacattagttccagtagcagcttcc  
ctcaacaagctcgtcttcggccagttactctcaggcatcagcttcttcgagcagcgggtg  
ccggatacacaggaccttctggaccttccactggaccgtctggctaccctgggcctttg  
agtggcggagcgtcgttcggctctggccaatcttctttcgggtcaaacttcagccttttc  
cgcatctgggtgctggacaatcggtcggagtatctgttatatcttctcttaattcacccg  
ttggattgaggtctccttctgctgcttctagacttagtcaattaacatcatccataacg

aatgcagttggtgccaatggtggtgatgctaattctcttgccgtagtcttcaatctag  
tttctcggcactcagaagctccggcatgtcttcaagcgatgctaaaattgaagtattgt  
tggaaccattggttggtctgcttcagcttttgagcaacactcaagtccgaggagtaa  
ccggcaacggcttcttcagtagcaaattctgctgagatcttttgaattagtttggc  
ttaagagatattgattgtagacctggagataaatgtaacttttctgatatgcaattg  
catacgaaatttcttattaaataaaagcattttgaaacattaaaaaaaaaaaaaaaa  
a

**ADDENDUM II**  
**Amino Acid Sequences**

SEQ ID NO: 29

Amino acid sequences encoded by SEQ ID NO: 1

AGQGGAGAAAAAAGGAGGAGRGGLGAGGAGQGYGSLGGQGGAGGGAAAAAAGG  
QGGQGGYGGLGSQGAGQGGYGAGQGGAGAAAAAAGGAGGAGRGGLGAGGAGQGYGSG  
LGGQGGAGQGGAAAAAAGGQGGQGGYGGLGSQGAGQGGAGRGAAAAAAGGQGGRG  
GYGGLGSQGAGQGGYGAGQGGAGAAAAAAGGAGGEGGLGAGGAGQGYGSLGGQGGAG  
QGGAAAAAAGGQGGHGGYGGLGSQGAGQGGAGRGAAAAAAGGQGGQGGYGGLGSQ  
GAGQGGYGAGQGGAAAAAAGGAGGAGRGELGAGGAGQGYGXGLGGQGGAGQRGAA  
SVAALAGGQGGQGGFGGFSSQGAGQGAYGGGAYSGQAAASVSAASAAASRLSSPGAAS  
RVSSAVTSLVSSGGPTNPAALSNTISXVVSQISE

SEQ ID NO: 30

Amino acid Sequences encoded by SEQ ID NO: 2

AAAAAAGGQGGQGGYDGLGSQGAGQGGYGQGGAAAAAASGAGSAQRGGLGAGGA  
QQGYGAGSGGQGGAGQGGAAAATAAAAGGQGGQGGYGGLGSQGSQGGYGQGGAAAAA  
AASGDGAGQEGLGAGGAGQGYGAGLGGQGGAGQGGAAAAAAGGQGGQGGYGGLGS  
QGAGQGGYGQGGAAAAAASGAGGAGQGGGAGAGQGYGAGSGGQGGAGQGGAAAAA  
AAAAGGQGGQGGYGGLGSQGAGQGGYGQGGVAAAAAASGAGGAGRGGLGAGGAGQGYG  
AVSGGQGGAGQGGAAAAAAGGQGGQGGYGGLGSQGAGQGGYGQGGAAAAAASGA  
GGARRGGLGAGGAGQGYGAGLGGQGGAGQGSASAAAAAAGGQGGQGGYGGLGSQGSQ  
GGYGQGGAAAAAASGAGGAGRGSLGAGGAGQGYGAGLGGQGGAGQGGAAAAASAAAG  
QGGQGGYGGLGSQGAGQGGYGQGGAAAAAASAGGQGGQGGYGGLGSQGAGQGGYGGGA  
FSGQGGGAASVATASAAASRLSSPGAASRVSSAVTSLVSSGGPTNSAALSNTISNVVSQ  
ISSNPNGLSGCDVLVQALLEIVSALVHILGSANIGQVNSSGVGRSASIVGQSINQAFS

SEQ ID NO: 31

Amino acid sequences encoded by SEQ ID NO: 3

AGSGQGGYGQGYGEGGAGQGGAGAAAAAAGGAGQGGQGGYGQGYQGGAGQGGAGAA  
AAAAAGGAGQGGYGRGGAGQGAAGAAAAAGSGQGGQGGYGQGYQGGAGQGGAGAAAAA  
AAGGAGQGGYGRGGAGQGGAAAAAAGGAGQGGQGGYGQGYQGGAGQGGAGAAAAA  
AGGAGQGGYGRGGAGQGGAAAAAAGGAGQGGYGRGGAGQGGAGSAAAAAAGGSGQGG  
QGGYGQGYQGGAGQGGAAAAASALAAPATSARISSHASTLLSNGPTNPASISNVISNAV  
SQISSNPNPGASSCDVLVQALLELVLTALLTIIGSSNVGNVNYDSSGQYAQVVSQSVQNAFV

SEQ ID NO: 32

Amino acid sequences encoded by SEQ ID NO: 4

GLGGQAGQAGAGAAAAAAGGAGQGGYGGLGSQGAGRGGYGGQAGAAAAAAGGAGQGGY  
GGLGSQGAGQGGYGGLGGQAGQGAAGAAAAAGGAGQGGYGGLGSQGAGRGGYGGQAGAA  
AAATGGAGQGGYGGVGSASASAAASRLSSPQASSRVSSAVSNLVASGPTNSAALSSTI  
SNAVSQIGASNPGLSGCDVLIQALLEVVVALIHILGSSSIGQVNYGSAGQATQ

SEQ ID NO: 33

Amino acid sequences encoded by SEQ ID NO: 5

GLGGQGAGRGAGAAAAAAGGAGQGGYGGLGGQGAGAAAAAAGGAGQGGQGLGGRGAAAAG  
GAGQGGYGGLGGQGAGRGAGAAAAAAGGAGQGGYGGLGGQGAGAAAAAAGGAGQGGYG  
GLGSQGAGRGGYGGQGAGAAVAAIGGVGQGGYGGVSGASASAAAASRLSSPEASSRVSS  
AVSNLVSSGPTNSAALSSTISNVVSQIGASNPGLSGCDVLIQALLEVVVSALVHILGSSSI  
GQVNYGSAGQATQ

SEQ ID NO: 34

Amino acid sequences encoded by SEQ ID NO: 6

SGLGGAGQGAGQGASAAAAAAXGGLGGGQGAGQGGQAGQGGYGSLGGAGQGASAAA  
AAAAAGGLGGGQGAGQGGQAGQGGYGSLGGAGQGASAAAAAAGGLGGGQGAGQGG  
QGAGQGGYGSLGGAGQGAGQGASAAAAAAGGLGGGQGAGYGSLGGVGGQGAGLGS  
RNSATNAISNASNAVSLSSPASNARISSAVSALASGAASGPGYLSSVISNVVSQVSSN  
SGGLVGCDTLVQALLEAAAALVHVLASSSGQVNLNTAGYTSQL

SEQ ID NO: 35

Amino acid sequences encoded by SEQ ID NO: 7

SGQGASAAAAAAGGLGGGQGAGYGSLGGAGQGGQAGQGAAAAAASAAAGGLGGGQGGQ  
QGAGRGGLQGAGQGGQAGLGGSRNSAANAVSRLSSPASNARISSAVSALASGGASSPGYL  
SSIISNVVSQVSSNNDGLSGCDTVVQALLEVAAAALVHVLASSNIGQVNLNTAGYTSQL

SEQ ID NO: 36

Amino acid sequences encoded by SEQ ID NO: 8

PGGAGQQGPGGQGPYPGAAAAAAGGYGPGAGQQGPXGAGQQGPGSQGPGGAGQQGPG  
GQGPYPGAAAAAAGGYGPGAGQQGPGSQGPGSGGQGGPGGQGPYPGPSAAAAAAGG  
YGPAGQQGPGSQGPGSGGQGGPGGLGYPGPSAAAAAAGGYGPGAGQQGPGSQGPGSG  
GQQRPGGLGYPGPSAAAAAAGGYGPGAGQQGPGSQGPGSGGQQRPGGLGYPGPSAAAA  
AAAAGGYGPGAGQQGPGSQAPVASAAASRLSSPQASSRVSSAVSTLVSSGPTNPAALSNA  
ISSVVSQVSASNPGLSGCDVLVQALLELVVSALVHILGSSSIGQINYAAS

SEQ ID NO: 37

Amino acid sequences encoded by SEQ ID NO: 9

AGPGYGPAGQQGPGSQGPGSGGQGGPGGQGPYPGPSAAAAAAGPGYGPAGQQGPGS  
GGQGGGQGSQGGPGGAGQGGPRGQGPYPGPSAAAAAAGGYGPGAGQQGPGSQGPGSG  
GQQGPGSQGPYPGPSAAAAAAGPGYGPAGQQGPGSQGPGSGGQGGPGGQGPYPGPSDA  
AAAAAAGPYGPAGQQGPGSGGQGGQGSQGGPGGAGQGGPRGQGPYPGPSAAAAAAG  
AGGYGPAGQQGPGSQGPGSGGQGGPGSQGPYPGPSAAAAAAGPGYGPAGQQGPGSQ  
GPGSGGQGGPGSQGPYPGPSAAAAAAGPGYGPAGQQGPGSQAPVASAAASRLSSPQA



SSRVSSAVSTLVSSGPTNPASLSNAISSVVSQVSSSNPGLSGCDVLVQALLEIVSALVH  
ILGSSSIGQINYAASSQYAQLVGQSLTQALG

SEQ ID NO: 38

Amino acid sequences encoded by SEQ ID NO: 10

GPGQQGPGGYGTSGPGGASAAAAAAGGPGGQGPSGPGPPGPGGYGPSGPGAAAAAA  
AAGGPGSQGPGQQGPGGYGPSGPGGASAAAAAAGGPGGQGSYGPQQGPGAGQYGP  
QQ

SEQ ID NO: 39

Amino acid sequences encoded by SEQ ID NO: 11

GQQGPGSQGPYPGPGAAAAAAGGYRPVSGQQGPGQQGPGSGGQQGPGGQRPYPGPGAA  
AAAAAGGYGPGSGQGGPGQQGPGSGGQQGPGGQGPYPGPGAAAAAAGGYGPGSGQGGQ  
QGPQSQGPSSGGQQGPGGQGPYPGSAAAAAAVGGYGPAGQQGPGQQGPGSGGQRP  
QGPYPGPGAAAAAAGGYGPASGQQGPGQGPSSGGQRPGGQGPYPGPGAAAAASAGGY  
GPGSGGSPASGAASRLSSPQAGARVSSAVSALVASGPTSPAASVSSAISNVASQISASNP  
LSGCDVLVQALLEIVSALVSILSSASIGQINYGASGQYAAMI

SEQ ID NO: 40

Amino acid sequences encoded by SEQ ID NO: 12

ASASGGAGPGRQQGYGPGGSGASAAAAAAGGAGPGGYGQGPSGYGPSGPGAQQGYGPG  
QGGSGAAAAAAGSGPGGYGPGAAGPGSYGPSGPGSGAAAAAAGSGPGGQQGYGPG  
GPGASAAAAAAGSGPGGYGQGPSGYGPSGPGAQQGYGPGGQGGSGAAAAAAGSGR  
GGYGPAAAGPBNYGPSGPGSGAAASAAAAAGPGGQQGYGPGGSGAAAAAAGGAGPGRQ  
QGYGPGGSGAAAAAAGSGPGGYGQGPAGYGPGGQGGSGAAAAAAGSGPGGYGP  
GAAGPBNYGPSGPGSGAAAAAAGSGPGGQQGYGPGGSGASAAAAAGGAGPGRQQAYGP  
GGSGAAAAAAGS

SEQ ID NO: 41

Amino acid sequences encoded by SEQ ID NO: 13

AGPGSYGPSGPGSGAAAAAAGSGPGGQQGYGPGGPGASAAAAAAGSGPGGYGQGPS  
GYGPSGPGAQQGYGPGGQGGSGAAAAAAGSGPGGYGPGAAGPBNYGPSGPGSGAAA  
SAAAAAGPGGQQGYGPGGSGAAAAAAGGAGPGRQQGYGPGGSGAAAAAAGSGSGPGG  
YGQGPXGYGPGGQGGSGAAAAAAGSGPXGYGPGAAGPBNYGPSGPGSGAAAAAAG  
SGPGGQQGYGPGGSGASAAAAAGGAGXGRQQAYGPGGSGAAAAAAGSGGYGPAQYXSS  
VASSAASAASALSSPTTHARISSHASTLLSSGPTNSAAISNVISNAVSQVSASNPSSSC  
DVLVQALLELITALISIVDSSNIGQVNYGSSGQYAQMVG

SEQ ID NO: 42

Amino acid sequences encoded by SEQ ID NO: 14

GSYGQGPSGYAQGSSAASAAAPSGYVPSQTGQSLGAAAAAAVAPSGYGPSQQGPSGP  
AATAAAGRPEGYGPRQQGPGATAAAGPGGYGPRQQGPGGYGPGQQGPGAAAAAAGR

GPGGYGPGQQGPGGPGAAAAAAGSEGYGPGQQGPRGPGAAAAAGPGGYGPGQQGASAAASA  
AAGRPGGYGPGQQGPGGPGSAAAAGPGGYGPGQQGPGSAAAAAAGSGPGGYGPGQQGPGG  
PGAAAAAAGPGGYGPGQQGPGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGRPGGYG  
PGQQGPGGPGAAAAAAGPGGYGPGGYGPGQQGPGGPGAAAAAAGRPGGYGPGQQGPGQ  
QPGGSGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGPGAAAAAAGR  
GPGGYGPGQQGPGGPGAAAAAAGRPGGYGPGQQGPGQQGPGGSGAAAAAAGRPGGYG  
PGQQGPGGPGAAAAAAGPGGYGPGQQGPGAAAAAAGRPGGYGPGQQGPGGSGAAAAA  
AGRPGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGTGAAAAAAGSGAGGYGPGQQGPG  
GPGAAAAAAGPGGYGPGQQGPGAAAAAAGSGPGGYGPGQQGPGGSSAAAAAAGPGRYGP  
GQQGPGAAAAASAGRPGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGPGAAAAAAGSG  
PGGYGPGQQGPGGPGAAAAAAGRPGGYGQGQQGPGGPGAAAAAAGPGGYGPGQQGPGA  
AAAAAAGSGPGGYGPGQQGPRSGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGPGG  
YGPQQGPGAAAAASAGRPGGYGPGQQGPGGSGAAAAAAGRPGGYGPGQQGPGGPGAA  
AAAAAGRPGGYGPGQQGPGQQGPGGSGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGP  
GGYGPGQQGPGAAAAAAGSGPGGYGPGQQGPGGPGAAAAAAGRPGGYGPGQQGPGGPG  
AAAAAAGRPGGYGPGQQGPGQQGPGGSGAAAAAAGRPGGYGPGQQGPGGPGAAAAA  
GPGGYGPGQQGPGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGRPGGYGPGQQGP  
GQQGPGGPGAAAAAAGPGGYGPGQQGTGAAAAAAGSGAGGYGPGQQGPGGPGAAAAAAG  
PGGYGPGQQGPGAAAAAAGSGPGGYGPGQQGPGGSSAAAAAAGPGRYGPQQGPGAAAA  
AAAGSGPGGYGPGQQGPGGPGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGPGGYGP  
GQQGPGAAAAAAGSGPGGYGPGQQGPGGPGAAAAAAGRPGGYGQGQQGPGGPGAAAA  
AAGPGGYGPGQQGPGAAAAAAGSGPGGYGPGQQGPRSGAAAAAAGRPGGYGPGQQ  
GPGGPGAAAAAAGPGGYGPGQQGPGAAAAASAGRPGGYGPGQQGPGGSGAAAAAAGRPG  
GGYGPGQQGPGGPGAAAAAAGTGPGGYGPGQQGPGGSGAAAAAAGRPGGYGPGQQGPG  
GPGAAAAAAGPGGYGPGQQGPGAAAAAAGSGPGGYGPGQQGPGGPGAAAAAAGRPGGY  
GPGQQGPGGPGAAAAAAGRPGGYGPGQQGPGGSGAAAAASGRPGGYGPGQQGPGGPG  
AAAAAAGPGGYGPGQQGPGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGRPGGYG  
PGQQGPGGSGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGPGAAAAAAG  
RPGGYGPGQQGPGGSGAAAAAAGRPGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGTG  
AAAAAAGSGAGGYGPGQQGPGGPGAAAAAAGPGGYGPGQQGPGAAAAAAGSGPGGYGP  
GQQGPGGPGAAAAAAGRPGGYGPGQQGPGGS

SEQ ID NO: 43

Amino acid sequences encoded by SEQ ID NO: 15

QGPSGPGSAAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPGSAAAAAAGPG  
QQGPGGYGPGPQGPGGYGPGQQGPGSYGPGQQGPGSGPGSAAASAAAAAGSQGPGGYGPG  
QQGPGGYGPGQQGPGSGPGSAAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPSAA  
AAAAAAGPGQQGPGGYGPGQQGPGGYGPGQQGPGSGPGSAAAAAAGPGQQGPGGYGPGQQ  
GPGGYGPGQQGPGSGPGSAAAAAAGPGQQGPGGYGPGQQGPGQQGPGSGPGSAAAAAAG  
GPGPQGPGGYGPGQQGPGGYGPGSGPGSAAAAAAGPGQQGPGGYGPGQQRPSGYGPGQQ  
GPSGPGSAAAAAAGPGQQGPGAYGPSGPGSAAAAAGLGGYGPAQQGPGSAGSAAAAAAG  
GPGGYGPVQQGPGSGPSAAGPGGYGPAQQGPARYGPGSAAAAAAGSAGYGPQPASAA  
ASRLASPDGARVASAVSNLVSSGPTSSAALSSVISNAVSQIGASNPGLSGCDVLIQALL  
EIVSACVTILSSSSIGQVNYGAA

SEQ ID NO: 44

Amino acid sequences encoded by SEQ ID NO: 16

QGPGGYGPSGPGSAAAAASAAAGPGQQGPGAYGPSGPGSAAAAAGPGXYGPGQQGPSGPGA  
AAAAAGPGQQGPGGYGPGAAAAAAAAGPGQQGPVAYGPSGPGSAASAAGPGGYGPARYG  
PSGSAAAAAAGAGSAGYGP GPQASAAASRLASPD SGARVASAVSNLVSSGPTSSAALSS  
VIXNAVSIQASNPGLSGCDVLIXALLEIVSACVTILSSSSIGQVNYGAA

SEQ ID NO: 45

Amino acid sequences encoded by SEQ ID NO: 17

AGGPGAGGAGAGGVPGGFGGPGGFGGAGGPGGPGGPGGAGGGAGGAGGLYGGGAGGL  
YGPGGLYGPGGAGVPGAPGASGRAGGIGGAAGGAGAGGVPGGVSSGAGGAGGSGVTVV  
ESVSVGGAGGPGAGGVPGGVPGGVPGGIYGPAGGAGGLYGPAGGAFFGPGGGAGAPG  
GPGGPGGPGGLGGGVGGAGTGGGVPGAGGVGPSGGAGGTGPVSVSSTVSVGGAGG  
PGAGGPGAGGAGGVPGGFGGPGGFGGAGGPGGPGGPGGAGGGAGGAGGLYGGGAG  
GLYGPGGLYGPGGAGVPGAPGASGRAGGIGGAAGAGGVPGGVSSGAGGSGVSVTESVT  
VGGAGGAGAGGIGGPSGLGGAGATGGFGGRGGPGGPGGPGGPRFGGAAGGAGAGGVGP  
GGVSSGAGGAGGSGVTVVESVSVGGAGGPGAGGVPGGVPGGVPGGIYGPAGGAGGLY  
GPGAGGAFFGSGGGAGAPGGPGGPGGPGGLGGGVGGAGTGGGVPGVGGVGPSGGAG  
GTGPVSVSSTITVGGGQSSGGVLPSTSYAPTTSGYERLPNLINGIKSSMQGGGFNYQNF  
GNILSQYATGSGTCNYYDINLLMDALLAALHTLNYQGASYVPSYSPSEMLSYTENVRR  
YF

SEQ ID NO: 46

Amino acid sequences encoded by SEQ ID NO: 18

GAPGGGPGGAGPGGAGFGPGGGAGFGPGGAAGGPGGPGGPGGAGGYGPG  
GAGGYGPGGVPGGAGGYGPGGAGGYGPGGSGPGGAGPGGAGGEGPVTVDVDVTVGPEGV  
GGGPGGAGPGGAGFGPGGGAGFGPGGAPGAPGGPGGPGGPGGPGGVPGGAGGYGPG  
GAGGVGPAGTGGFGPGGAGGFGPGGAGGFGPGGAGGFGPGGAGGYGPGGVPGGAGGFGP  
GGVPGGSGPGGAGGEGPVTVDVDVSVGGAPGGGPGGAGPGGAGFGPGGGAGFGPGGGAG  
FGPGGAAGGPGGPGGPGGPGGAGGYGPGGAGGYGPGGVPGGAGGYGPGGAGGYGPGGSG  
PGGAGPGGAGGEGPVTVDVDVTVGPEGVGGGPGGAGPGGAGFGPGGGAGFGPGGAPGAPG  
GPGGPGGPGGPGGPGGVPGGAGGYGPGGAGGVGPAGTGGFGPGGAGGFGPGGAGGFGPG  
GAGGFGPAGAGGYGPGGVPGGAGGFGPGGVPGGSGPGGAGGEGPVTVDVDVSVGGAPG  
GGPGGAGPGGAGFGPGGGAGFGPGGGAGFGPGGAAGGPGGPGGPGGPGGAGGYGPGGAGG  
YGPGGVPGGAGGYGPGGAGGYGPGGSGPGGAGPGGAGGEGPVTVDVDVTVGPEGVGGGPG  
GGAGPGGAGFGPGGGAGFGPGGAPGAPGGPGGPGGPGGPGGPGGVPGGAGGYGPGGAGG  
FGPGGTGGFGPGGAGGFGPGGAGGFGPGGAGGFGPGGAGGYGPGGVPGGAGGFGPGGVG  
PGGSGPGGAGGEGPVTVDVDVSVGGAPGGGPGGAGPGGAGFGPGGGAGFGPGGGAGFGPG  
GAAGGPSGPGGPGGPGGAGGYGPGGAGGYGPGGVPGGAGGYGPGGAGGYGPGGSGPGGA  
GPGGAGGEGPVTVDVDVTVGPEGVGGGPGGAGPGGAGFGPGGGAGFGPGGAPGAPGGPGG  
PGGPGGPGGPGGVPGGAGGYGPGGAGGVGPAGTGGFGPGGA

SEQ ID NO: 47

Amino acid sequences encoded by SEQ ID NO: 19

GSGQGGRYGGQSSGGYGQGAGAGAATAATARADGSGQGRYDGQSSQGGYGQGAGAGATA  
TAAAGGAGSGQGGYGGQGLGGYGQGAGAGAAAAATAAGGAGSGQGDYGDQGLGGYGQG  
SGAGSATAPAAGGSGFGQGGFGRNRGGKAYQGSAGAGVGAATAAAGGAGSGQGGYGDQ  
GGLGGYGQGAGAGAASAAAGGDDGYEQGGYGNQGLGSFGQAGAGAAAAASAGGAGSG  
RGGYGDQGLGGYGQGAGAGAASAAAGGDDGYGQGYGDQGGRRGGYGQSGAGSATAAA  
AGGAGFGQGGYGQGGYGNQGLGSFGQAGAGAAAAASAGGAGSGRGGYGDQGLGGYG  
QGAGAGAAAAAAGGDDGYGQGGYGNQGLGSFGQAGAGAAAAASAGGAGSGRGGYQG  
GYGNQGLGSFGQAGAGAAAAASAGGAGSGRGGYGDQGLGGYGQGAGSGAAAAAAGG  
GDGYGQGGYGNQGLGSFGQAGAGAAAAASAGGAGSGRGGYQGGYGNQGLGSFGQG  
AGAGAAAAASAGGAGSGRGGYGDQGLGGYGQGAGAGAASAAAGGDDGYGQGGYGNQRG  
VGSYGQGAGAGAAATSAAGGAGSGRGGYGEQGLGGYGQGAGAGAAATAAGGDDGYQG  
GYGNQGGRGSYQGSGAGAGAAVAAAAGGAVSGQGGYDGEQGQGGYGQSGAGAAVAAA  
SGGTGAGQGGYGSQGSQAGYGQAGFRAAAATAAAGAGGAGGGQGGYGGQGGYGQGTGA  
GGASSAGLSVTGNMVSRLSSPEAASRVSSAVSSLSVNGQVNVDALPSIIISNLSSSISA  
SATTASDCEVLVQVLLLEVVSALVQIVCS

SEQ ID NO: 48

Amino acid sequences encoded by SEQ ID NO: 20

YGQSGAGAAAAAAAGGAGQSGSGPYGASYLSSTTYTTSSQAGGGVGGYGQSGTGSA  
AAAAGAAGAGQGGQGGYGQGAGQGLGGYGQGGGAGAAAAAAAGGAGSGQGGYGGQGG  
LGGYGQGAGAGAAAAAAAGGAGAGQGGFGGQGGYGQGGGAGAAAAAAAGGAGSGQG  
GYGGQGLGGYGQGAGAGAAAAAAAGGAGAGQGSYGGQGGYGQGGAGAAATAAAAGGAG  
SGQGGYGGQGLGGYGQGAGAGAAAAAAAGGAGAGQGGYGGQGGQGGYGQGAGAGAAA  
AAAGGAGAGQGGYGGQGGYGQGGGAGAAAAAAASGGSGSGQGGYGGQGLGGYGQGAGA  
GAGAAASAAAGAGSGQGGYGGQGLGGYGQGAGAGAAAGASGSGSGGAGQGLGGYGQG  
AGAGAAAAAGASGAGQGGFGPYGSSYSSTYSVTSQGAAGGLGGYGQSGAGAAAAAGA  
AGQGGQGGYGQGAGAGAGAGQGLGGYGQGAGSSAASAAAGGAGAGQGGYGGQGLG  
GYQGAGAGASAAASASGAGSGQGGYGGQGGYGQGTGAGAASSAGVAVTVGNTVSRSSP  
QAASRVSSAVSSLSVNGQVNVAALPSIIISLSSSISASSTAASDCEVLVQVLLIVSALV  
QIVSSANVGYNPEASGSLNAVGSALAAAMG

SEQ ID NO: 49

Amino acid sequences encoded by SEQ ID NO: 21

NASQIAASVASAVASSASAAAAAASSSAAAAAGASSAAGAASSSSTTTTTSTSSSAAA  
AAAAAASASGASSASAAASASAAAFSSALISDLLGIGVFGNTFGSIGSASAASSIA  
SAAAQAALSGLSYLASAGASAVASAVAGVGVGAGAYAYAYAIANAFASILANTGLLS  
VSSAASVASSVASAIATSVSSSSAAAAASASAAAAASASAASSASASSSASAAAAAGAS  
AAAGAASSASASAAAFSSAFISALLGFSQFNSVFGSITSASLGLGIAANAVQSGLAS  
LGLGAAASAAASAVANAGLNGSGAYAYATAIASAIGNALLGAGFLTAGNASQIAASVAS  
AVASSASAAAAAASSSAAAAAGASSAAGAASSSSTTTTTSTSSSAAAAAASASGA  
SSASAAASASAAAFSSALIGDLLGIGVFGNTFGSIGSASAASSIASAAAQAALSGLG  
LSYLASAGASAVASAVAGVGVGAGAYAYAYAIANAFASILANTGLLSVSSAASVASSVA

SAIATSVSSSSAAAAASASAAAAASAGASAASSASASSSSASAAAGAGAGAGAGASGASG  
AAGGSGGFGGLSSGFGAGIGGLGGYPSGALGGLGIPSGLLSSGLLSPAANQRIASLIPLI  
LSAISPNGVNFVIGSNIASLASQISQSGGGIAASQAFTQALLELVAAFIQVLSSAQIG  
AVSSSSASAGATANAFQSLSSAFAG

SEQ ID NO: 50

Amino acid sequence encoded by SEQ ID NO: 22

AAAAAAAAAAGAGAGAGAGAGAGAGAGAGSGASTSVSTSSSSGSGAGAGAGSGAGSGAGAG  
SGAGAGAGAGGAGAGFGSGLGLGYGVGLSSAQQAQAQAAAQAQAQAQAYAAAQAQA  
QAQAQAQAAAAAAGAGAGAGAGAGAGAGAGSGASTSVSTSSSSGSGAGAGAGSG  
AGSGAGAGSGAGAGAGAGGAGAGFGSGLGLGYGVGLSSAQQAQAQAAAQAQAQAQA  
YAAAQAQAQAQAQAAAAAAGAGAGAGAGAGAGAGAGSGASTSVSTSSSSGSG  
AGAGAGSGAGSGAGAGAGAGAGAGAGGAGAAFGSGLGLGYGVGLSSAQQAQAQAAAQA  
QADAQAQAYAAAQAQAQAQAQAAAAAAGAGAGAGAGSGAGAGAGSGASTSVS  
TSSSSGSGAGAGAGSGAGSGAGAGSGAGAGAGAGGAGAGFGSGLGLGYGVGLSSAQQA  
QAQAAAQAQADAQAQAYAAAQAQAQAQAQAAAAAAGAGAGAGAGSGAGAGAG  
SGASTSVSTSSSSGSGAGAGAGSGAGSGAGAGSGAGAGAGAGGAGAGFGSGLGLGYGV  
LSSAQQAQASAAAQAQADAQAQAYAAAQAQAQAQAQAAAAAAGAGAGAGAG  
AGAGAGAGSGASTSVSTSSSSASGAGAGAGSGAGSGAGAGSGAGAGAGAGGAGAGFGSG  
LGLGYGVGLSSAQQAQAQAAAQAQAQAQALAAAQAQAQAQAQAAAAATAAAAAAG  
AGAGAGSGAGAGAGAGAGSGASTSVSTSSSSAAGAGAGAGSGAGAGSGTGAGIALPSIV  
LSPAASSRISVSSSVQSAGSGLSFSSLSNTLSQTASAIRSSNPQLSSSDVLIQSLVEI  
IVGLVQAFTGSSASAQTFVNSLSQVAG

SEQ ID NO: 51

Amino acid sequence encoded by SEQ ID NO: 23

TDSVASSASSSASASSSATGPDGTGYPVGYGAGQAEAAASAAAAAASAAEAATIAGLGY  
GRQGQGTDSASSVSTSTSVSSLATGPGSRYPVVDYADQAEAAASAAAAAASAAEIASL  
GYGRQGQGTDSVASSASSSASASSSATGPDGTGYPVGYGAGQAEAAASAAAAAASAAEA  
ATIAGLGYGRQGQGTDSASSVSTSTSVSSSATGPDGTGYPVGYGAGQAEAAASAAAAA  
ASAAEAATIAGLGYGRQGQGTDSASSVSTSTSVSSSATGPDGMGYPVGNYGAGQAEAAAS  
AAAAAASAAEAATIASLGYGRQGQGTDSASSVSTSTSVSSSATGPGSRYPVVDYADQ  
AEAAASAAAAAASAAEIASLGYGRQGQGTDSVASSASSSASASSSATGPDGTGYPVG  
YYGAGQAEAAASAAAAAASAAEAATIAGLGYGRQGQGTDSASSVSTSTSVSSSATGPG  
SRYPVVDYADQAEAAASATAAAAAAASAAEIASLGYGRQGQGTDSVASSASSSASASS  
SATGPDGTGYPVGYGAGQAEAAASAAAAAASAAEAATIAGLGYGRQGQGTDSASSVST  
STSVSSSATGPGSRYPVMDYADQAEAAASAAAAAASAAEAATIAGLDYEGQGQGTDSGASS  
VSSSTSVSSSATGVTQTIALPPDVSARISFLTSYLQSAGSGLSLYTLNLLSQTALAIS  
KSRPELSPNEVLIQSLAEIIVALVQALTKQASSSASVQYFGRFL

SEQ ID NO: 52

Amino acid sequence encoded by SEQ ID NO: 24

AISSSLYAFNYQASAASSAAQSSAQTAASTSAKQTAASTSASTAATSTTQTAATTASTA  
ASSQTQKASTSSAASTAASKSQSSSAGSSRTTSTAASASSSYAFAQSLSQYLLSSQQF

TTAFASSTAVASSQQYAEAMAQSVATSLGLGYTYASALSVAMAQAIISGVGGGASAYSAT  
AISQAIISRALTSSGVSLSSSQATSVASAISSSLYAFNYQASAASSAAQSSAQTAASAK  
QTAASTSASTAATSTTQTAATTSASTAASSQTVQKASTSSAASTAASKSQSSSVGSSTTS  
TAAASASSSYAFAQSLSQYLLSSQQFTTAFASSTAVASSQQYAEAMAQSVATSLGLGYTY  
TSALSVAMAQAIISGVGGGASAYSATYATISQAIISRVLTSSGISLSSSQATSVASAISSSLY  
AFNYQASAASSAAQSSAQTAASAKQTAASTSASTAATSTTQTAATTSASTAASSQTVQ  
KASTSSAASTAASKSQSSSVGSSTTSTAASASSSYAFAQSLSQYLLSSQQFTTAFASST  
AVASSQQYAEAMAQSVATSLGLGYTYTSALSVAMAQAIISGVGGGASAYSATYATISQAIIS  
VLTSSGVSLSSSQATSVASAISSSLYAFNYRASAASSAAQSSAQTAASAKQTAASTSA  
STAATSTTQTAATTSASTAASSQTVQKASTSSAASTAAQQTGQSSSVQNQGSSSASSSSV  
SVSDISDSLTTSLQSEFTSAFGSTVSEAEAQSYAEVAQSTVAQLGIDYSQSSALATA  
VANAVSQVKQGSSSRAYARAIYAITTYLKTTRIITITRTQVKSFASAISSSLSTARAT  
SSANAYQEQTQSSAAASAAQSSSEYQTQNTQSSASAASSDASTSYQTQSSYSDASAASV  
AAESTSANQAQSTQSSAAASSSTNSAYQSQSYIDASTVSSASANTAQSTYQVTIPDNTY  
FAESLSSTLIQHEQFNSKFGSYIPLVTAREYASAMARATALIIGFDSTGTSALSAVAVA  
VSNVDYASAYSARAIYAFINVLNNGIFASASEALYLAPAMIASLHAFGKSSFSESSA  
FALANSISPSTAITSAQSSSVSAGASSGQSSYDTSSVSSASSAEATESSSVFDTYQATQ  
IESSAAAAAASSSAYDSQFSESSSASSAAASAFSEQTSYDISSDLSSASSATAAAASSSA  
YESQFSDASSGSSAAAAASSQONSQYDTDALYSASSAASAAASASAYELEFSDASSSSSAV  
AVASSQQGSYDTSSDFSSASSAAAAASAYESKFLDASSSSSAAAAASSQSSYETSSDL  
VSASSAAAAASASAYQSQFLDASSSSNAAATTSSRQSSYDTSSDFSSASIAAAAAASAS  
SYESQFSDASSSSNAAAAASSQSSYDTSSDLVSAASASAYESQFLDASSSSNAAATTSS  
QSSYDTSSDFSSASIAAAAAASASSYESQFSDASSSSNAAAAASSQSSYDTSSDLVSA  
SSAAAAAASASSYESQFSDASSSSNAAAAASSQSSYDTSSDLVSAASAAAAASASAYE  
SQFSDASSSRNAAAAASSQSSYDTSSDLVSAASAAAAAASASSYESQFLDASSSSNAAA  
TTSSQSSYDTSSDFSSASIAAAVAASASSYESQFSDASSSSKAAAAASSQSSYDTSSD  
LVSAASAAAAAASASSYESQFSDASSSSNAAAAASSQSSYDTSSDFSSANSAAAESSA  
ATEIYQETQIASSIAAASALSEHTSELAEESSSSSAASAAAAAASEQSLYDTSSAASSA  
SSSDFIASSDIRNQQLSVNSAASSSAAEESVSQVDEETYQNFQYSSISASASAAQSS  
IYQDVSSSSAAASTSSAASSLETSGTVAESGSTAASSSYAAAAAASSSAGSTSSPSFLSA  
DSLSSSLASLRICSFSSKLMSSLYSGDGLDIAEFSDAVSSMVSSIKSSNPGVSASQILTE  
LLFEVIVAFVQALTKSKFSTMETAESLIAAFAQAFV

SEQ ID NO: 53

Amino acid sequence encoded by SEQ ID NO: 25

SQQGPIGGVGGSNAFSSSFASALSLNRGFTEVISSASATAVASAFQKGLAPYGTAFALSA  
ASAAADAYNSIGSGANAFAYAQAFARVLYPLVRQYGLSSSGKASAFASAIASSFSSGTSG  
QGPGSIGQQPPVTISAASASAGASAAAVGGGQVGGQGPYGGQQQSTAASASAAAATATSGA  
AQKQPSGESSVATASAAATSVTSGGAPVGKPGVPAPIFYPPQGPLQQGPAPGPSNVQPPTS  
QQGPIGGVGGSNAFSSSFASALSLNRGFTEVISSASATAVASAFQKGLAPYGTAFALSAA  
SAAADAYNSIGSGANAFAYAQAFARVLYPLVRQYGLSSSGKASAFASAIASSFSSGTSGQ  
GPSIGQQQPPVTISAASASAGASAAAVGGGQVGGQGPYGGQQQSTAASASAAAATATSGAA  
QKQPSGESSVATASAAATSVTSGGAPVGKPGVPAPIFYPPQGPLQQGPAPGPSNVQPPTSQ  
QGPIGGVGGSNAFSSSFASALSLNRGFTEVISSASATAVASAFQKGLAPYGTAFALSAAS  
AAADAYNSIGSGANAFAYAQAFARVLYPLVQQYGLSSSAKASAFASAIASSFSSGTSGQG  
PSIGQQQPPVTISAASASAGASAAAVGGGQVGGQGPYGGQQQSTAASASAAAATATSGGAQ

KQPSGESSVATASAAATSVTSAGAPVGKPGVPAPIFYPPQGPLQQGPAPGPSNVQPGTSQQ  
GPIGGVGGSNFSSSFASALSLNRGFTEVISSASATAVASAFQKGLAPYGTAFALSAASA  
AADAYNSIGSGANAFAYAQAFAFARVLYPLVQQYGLSSSAKASAFASAIASSFSSGTSGQGP  
SNGQQQPPVTISAASASAGASAAAVGGGQVSQGPYGGQQQSTAASASAAAATATSGGAQK  
QPSGESSVATASAAATSVTSAGAPGGKPGVPAPIFYPPQGPLQQGPAPGPSNVQPGTSQQG  
PIGGVGGSNFSSSFASALSLNRGFTEVISSASATAVASAFQKGLAPYGTAFALSAASAA  
ADAYNSIGSGANAFAYAQAFAFARVLYPLVQQYGLSSSAKASAFASAIASSFSSGTSGQGPS  
IGQQQPPVTISAASASAGASAAAVGGGQVGQGPYGGQQQSTAASASAAAATATSGGAQKQ  
PSGESSVATASAAATSVTSAGAPVGKPGVPAPIFYPPQGPLQQGPAPGPSYVQPATSQQGP  
IGGAGRSNAFSSSFASALSGNRGFSEVISSASATAVASAFQKGLAPYGTAFALSAASAAA  
DAYNSIGSGANAFAYAQAFAFARVLYPLVQQYGLSSSAKASAFASAIASSFSSGAAGQQQSI  
PYGGQQQPPMTISAASASAGASAAAVKGGQVGQGPYGGQQQSTAASASAAATTATAGGAQ  
KHPSGEYSVATASAAATSVTSGGAPVGKPGVPAPIFYPPQGPLQQGPAPGPSNVQPGTSQQ  
GPIGGVGESNTFSSSFASALGGNRGFSGVISSASATAVASAFQKGLAPYGTAFALSAASA  
AADAYNSIGSGASASAYAQAFAFARVLYPLLQQYGLSSSADASAFASAIASSFSTGVAGQGP  
SVPYVGGQQPSIMVSAASASAAAASAAVGGGPVVQGPYDGGQPQQPNIAASAAAAATATS  
SGPKEEPLGESSVIATSVSAASSVSSGGAPGVQGGGPVTVSYREGPSQIPSQQTLLQAVP  
STQSVGSGVPVGPNQYEMVYAPLQQFGGVSASNLLSPSAHSRIASLMSDVLSLFSFGNSG  
FNYGGFARALSSVARAVSQSNAKLSTTDVIIQVLMEALVALIELLSGAKIGVVHPVRAQA  
GASAFQAQHFQSAFG

SEQ ID NO: 54

Amino acid sequences encoded by SEQ ID NO: 26

AGAGAGYGAGAGSGAGAGSGAGAGAGAGAGAGAGYGAGAGSGAGAGAGYGRGAGAGAGA  
GAGYGQGAGAGAGAGAGAGYGAGAGSGAGAGYGTGAGSGAGSGAGSGAGAGAGAGAG  
YGAGAGSGAGAGAGYGRGAGAGAGAGAGAGYGQGAGAGAGAGAGAGSGAGAGSGAGAGAGA  
GYQGAGAGAGAGAGAGYGAGAGSGAGAGAGYGAGAGSGAGAGSGAGAGAGAGYGQGAGAG  
AGAGYGAGAGSGAGAGSGAGAGSGAGAGSGYGAGAGSGAGAGAGYGQGAGAGAGAGAGY  
GAGAGSGAGTGAGYGAGAGAGYGAGAGAGAGSGAGAGAGYGAGAGAGAGAGYGAGAGSG  
XGAGXGYGAGAGAGSGVGAGAGAGAGAGYGAGAGAGAGYGAGAGAGAGAGAGAGAGYGA  
GAGASVSSSTVSNTASRMSSSENTSRRVSSAISSIVGSGGVNMNSLSNVISNVSSSVAASN  
PGLSGCEVLVQTLLEVVSALVHILSYASVGSVDASAAGQSAQTVATAMSSVMG\*

SEQ ID NO: 55

Amino acid sequence encoded by SEQ ID NO: 27

GAAAAASAAAAGGRGSQGGYGDDGGAIAAAAAAAAAAAGSGGTGGGQGGRGDGGAAAAA  
AAAAEAAAGGKGRQGSYGDDGGAIAAAAAAAAAAAGRGGSGRGQLRRDKGSYGVDGGAE  
AAASAAATAGRQGRQGSYGDDGGAIAAAAAAAAAASASRLASSSAVSRVSSAVSALLSNGFSD  
VNSLSNVISGLSASVSSSTPELTGCEVLVEVLLEVVSALVHILNFADIGNVNISASGDST  
SLVGRTVLEAFG\*

SEQ ID NO: 56

Amino acid sequences encoded by SEQ ID NO: 28

GINVDSGSVQSDISSSSSFLSTSSSSASYSQASASSSSGAGYTGPSGPSTGPSGYPGPL  
SGGASFGSGQSSFGQTSAFSASGAGQSAGVSVISSLNSPVGLRSPSAASRLSQLTSSIT  
NAVGANVDANSLARSLQSSFSA LRSSGMSSSDAKIEVLLETIVGLLQLLSNTQVRGVN  
PATASSVANSAARSFELVLA\*

SEQ ID NO: 57

Consensus amino acid sequence of a 200 amino acid repeat  
unit of SEQ ID NO: 56

SSVVQRAAQSLASTLGVDGNNLARFAVQAVSRLPAGSDTSAYAQAFSSALFNAGVLNAS  
NIDTLGSRVLSALLNGVSSAAQGLGINVDSGSVQSDISSSSSFLSTSSSSASYSQASAS  
STSGAGYTGPSGPSTGPSGYPGPLGGGAPFGQSGFGGSAGPQGGFGATGGASAGLISRV  
ANALANTSTLRTVLRTGVSQQIA